

Ribonucleolytic resection is required for repair of strand displaced nonhomologous end-joining intermediates

Article (Accepted Version)

Bartlett, Edward, Brissett, Nigel C and Doherty, Aidan J (2013) Ribonucleolytic resection is required for repair of strand displaced nonhomologous end-joining intermediates. *Proceedings of the National Academy of Sciences*, 110 (22). E1984-E1991. ISSN 1091-6490

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/47178/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Ribonucleolytic resection is required for repair of strand displaced NHEJ intermediates

Edward Bartlett, Nigel C. Brissett & Aidan J. Doherty

Genome Damage and Stability Centre, School of Life Sciences, University of Sussex,
Brighton BN1 9RQ, UK

§ Corresponding author: email: ajd21@sussex.ac.uk

Keywords: Double-strand break; DNA; repair; NHEJ; Archaea; RNA

Abstract

Non homologous end-joining (NHEJ) pathways repair DNA double-strand breaks (DSBs) in eukaryotes and many prokaryotes, although not reported to operate in the third domain of life, archaea. Here, we describe a complete NHEJ complex, consisting of DNA ligase (Lig), polymerase (Pol), phosphoesterase (PE) and Ku from a mesophilic archaeon, *Methanocella paludicola* (*Mpa*). *Mpa* ligase has limited DNA nick sealing activity but is efficient at ligating nicks containing a 3' ribonucleotide. *Mpa* Pol preferentially incorporates NTPs onto a DNA primer strand, filling DNA gaps in annealed breaks. *Mpa* PE sequentially removes 3'-phosphates and ribonucleotides from primer strands, leaving a ligatable terminal 3'-monoribonucleotide. These proteins, together with the DNA end-binding protein Ku, form a functional NHEJ break repair apparatus that is highly homologous to the bacterial complex. Although the major roles of Pol and Lig in break repair have been reported, PE's function in NHEJ has remained obscure. We establish that PE is required for ribonucleolytic resection of RNA intermediates at annealed DSBs. Polymerase-catalyzed strand displacement synthesis on DNA gaps can result in the formation of non-ligatable NHEJ intermediates. PE's function in NHEJ repair is to detect and remove inappropriately incorporated ribonucleotides, or phosphates, from 3' ends of annealed DSBs to configure the termini for ligation. Thus, PE prevents the accumulation of abortive genotoxic DNA intermediates, arising from strand displacement synthesis, that would otherwise be refractory to repair.

NHEJ is a major DNA break repair pathway in eukaryotes and prokaryotes but assumed to be absent in archaea. This study establishes that a functionally homologous pathway is present in archaea. We have reconstituted archaeal NHEJ repair *in vitro*, demonstrating that it is closely related to the bacterial apparatus and preferentially repairs breaks using RNA intermediates. We identify a role for a functionally unassigned nuclease in preventing the accumulation of genotoxic repair intermediates produced by strand displacement. This study has important implications for our understanding of the mechanisms of DNA break repair by NHEJ and the evolution of end-joining pathways.

\body

Introduction

DNA double-strand breaks (DSBs) are one of the most lethal forms of damage encountered by cells (1). Non homologous end-joining (NHEJ) is the primary pathway for repairing DSBs in higher eukaryotes (2). NHEJ does not require the presence of a sister chromatid, in contrast to homologous recombination (HR), and can therefore operate in quiescent cells. The direct repair and ligation of DSBs by NHEJ is considered to be more error-prone as a consequence of mending breaks without the assistance of an intact DNA template.

Higher eukaryotic NHEJ complex is primarily composed of Ligase IV, XRCC4, XLF (LXX complex), DNA-PKcs, Ku70/80 (1, 3). Ku and DNA-PKcs both assist in bridging the gap between the broken termini by promoting end-synapsis. NHEJ repair of non homologous breaks requires remodelling of the DNA termini by processing enzymes, including polymerases (Pol λ and μ) and nucleases (Artemis and Fen-1), to prepare them for ligation (4-7). The LXX complex is recruited by Ku, which enables the ligation of the DNA ends.

Most bacterial organisms also possess a NHEJ complex, composed of a more limited set of proteins, including Ligase D (LigD) and Ku (8, 9), which is required for DSB repair in stationary phase (10). The mycobacterial LigD gene encodes a multifunctional enzyme which, along with Ku, is capable of remodelling and ligating a broad variety of DSBs (8, 11). In many bacteria, LigD is found as a fusion of DNA ligase (Lig), polymerase (Pol) and phosphoesterase (PE) domains, in a variety of configurations. The NHEJ ligase (Lig or LigDom) has a strong propensity for ligating nicks with a ribonucleotide on the 3' position of the break (12) and the polymerase (Pol or PolDom) preferentially incorporates NTPs, rather than the expected dNTPs (12, 13). Together, these findings establish that NHEJ repair requires polymerase insertion of ribonucleotides at the DNA termini for efficient ligation. The phosphoesterase (PE or NucDom) domain of LigD, originally shown to have 3' exonuclease activity (8), preferentially removes ribonucleotides and phosphates from the 3' termini strand, by sequential phosphodiesterase and phosphomonoesterase activities (14). Whilst PE's biochemical activities are known, the specific role of this nuclease in DSBs repair remains unclear, with the nucleolytic activities of PE appearing to be antagonistic to those of the polymerase. Indeed, no obvious alterations to repair efficiency or fidelity was observed in plasmid-based DSB repair assays that excluded PE (15).

Archaea make up the third major domain of life and, although these organisms are one of the most abundant in the biosphere, relatively little is known about how they repair their DNA. Many archaeal species survive in some of the most extreme environments on the planet and therefore it is of importance to investigate

how they maintain genome stability to understand how life evolved in the hostile conditions present on the primordial earth and how they continue to enable organisms to thrive in extreme terrestrial niches. Although some NHEJ-like genes have been identified in archaea (9, 16-18), a complete DNA break repair apparatus has not been reported. Therefore, it is unclear if a related DSB repair pathway operates in archaea. Here, we describe the discovery and characterization of a complete NHEJ complex, composed of Lig, Pol, PE and Ku, in the archaeon, *Methanocella paludicola* (*Mpa*). This study establishes, for the first time, that a functional NHEJ apparatus exists in archaea that is highly homologous to the bacterial break repair complex. Furthermore, we have exploited the biochemically tractability of this NHEJ apparatus to identify and characterize an unexpected function of the PE nuclease activity in co-ordinating end-joining. We demonstrate that the phosphoesterase is required for ribonucleolytic resection of polymerase-incorporated RNA intermediates, resulting from strand displacement synthesis, at annealed DSBs. This unusual 3' resection activity configures nicked NHEJ intermediates for optimal end-joining thus preventing the accumulation of potentially lethal non-ligatable intermediates that can arise from strand displacement synthesis by NHEJ polymerases.

RESULTS AND DISCUSSION

Identification of a putative NHEJ repair complex in archaea

NHEJ pathways operate in both eukaryotes and prokaryotes but the existence of a related DSB repair system exists in the third domain of life, archaea, is unclear. Although a few NHEJ-like genes have been identified in archaea, a complete repertoire of factors has not been discovered in any one species, suggestive that a functional end-joining repair pathway may not exist in this kingdom. To address this question, mycobacterial NHEJ protein sequences (Ku and LigD) were used to iteratively search for related sequences in archaeal genomes. We identified that the genome of a mesophilic archaeon, *Methanocella paludicola* (*Mpa*) (19), possesses an apparently intact NHEJ apparatus. *Mpa* encodes NHEJ ligase (Mcp 2126), polymerase (Mcp 2125) and phosphoesterase (Mcp 2127) genes, located in an operonic arrangement, that share significant homology with the three distinct catalytic domains of mycobacterial LigD (Fig. 1A; Fig. S1). However, unlike LigD, the different putative activities are encoded on individual genes rather than fused together into a single LigD-like gene. *Mpa* also encodes the DSB end-binding protein, Ku (Mcp 0581), a “hallmark” gene for the presence of a *bona fide* NHEJ pathway. Together, these *Mpa* repair protein orthologues represent the first example of a complete NHEJ complex identified in an archaeal organism.

Mpa NHEJ proteins possess equivalent activities of prokaryotic orthologues

To allow biochemical characterization of the putative *Mpa* NHEJ proteins, the genes were individually over-expressed in *E. coli*, fused to a N-terminal hexa-histidine tag. The proteins were purified to homogeneity, in high yields, using immobilized metal

affinity chromatography (IMAC), ion exchange chromatography and gel filtration chromatography (see methods).

Mpa Ku gene encodes a 252 amino acid polypeptide that shares significant sequence homology with the prokaryotic and viral Ku proteins. Following purification, *Mpa* Ku was observed to migrate at ~30kD by SDS-PAGE (Fig. S2) and eluted as an apparent homodimer from gel filtration columns, consistent with other Ku proteins (9, 20). Increasing titration of *Mpa* Ku reduced the electrophoretic mobility of double-stranded (ds) DNA (Fig. 1B), confirming that it binds to ds DNA, consistent with other Ku proteins (13).

Mpa DNA polymerase (*Mpa* Pol) gene encodes a 295 aa protein (~34kD; Fig. S2) containing the three Archaeo-Eukaryotic Primase (AEP) catalytic motifs present in other LigD-associated primase-polymerases (Fig. S3). In primer-template extension assays, *Mpa* Pol incorporated either dNTPs or NTPs, in the presence of manganese, opposite a DNA template of a 5' overhang (Fig. 1C). Similarly to other bacterial NHEJ polymerases (13), *Mpa* Pol displayed a marked preference for insertion of NTPs over dNTPs, with optimal activity observed in the presence of manganese ions, rather than magnesium (Fig. S2).

Mpa phosphoesterase (PE) gene encodes a 198 aa protein, migrating ~24kD in SDS-PAGE (Fig. S2). *Mpa* PE represents a full-length, stand-alone, orthologue of the LigD PE domain. *Mpa* PE cleaved 3'-ribonucleotides and 3'-phosphates from a recessed DNA-RNA primer strand, leaving a single 3' ribonucleotide containing a 3'-

hydroxyl group (Fig. 1E), in common with the orthologous ribonucleolytic activity reported for other 3'-phosphoesterases (14, 18). As previously reported, this ability to remove all but the last ribonucleotide from a primer strand indicates the necessity for the phosphoesterase to recognise a 2'-hydroxyl ribose moiety on the penultimate base. *Mpa* PE required either manganese or cobalt as a cofactor for efficient catalytic activities. Copper, cadmium and magnesium all elicited moderate ribonuclease and phosphatase activities (Fig. S4).

Mpa DNA ligase (*Mpa* Lig) gene encodes a 334 aa protein, which migrated at ~40kD in SDS-PAGE (Fig. S2). *Mpa* Lig possesses all of the conserved motifs found in members of the ligase nucleotidyltransferase superfamily (21, 22) (Fig. S5). Notably, *Mpa* Lig had limited ligation activity on nicked DNA substrates (Fig. S6). In contrast, it preferentially catalyzed the ligation of nicked DNA containing at least one ribonucleotide on the 3'-hydroxyl (3'-OH) side of a nick (Fig. 1C). Ligation was optimal in the presence of manganese, with only limited end-joining in the presence of magnesium (Fig. S6). *Mpa* Lig was able to catalyze the ligation of nicks with a 5'-phosphate and a 3'-monoribonucleotide in the absence of ATP, suggesting that some of the recombinant protein was pre-adenylated, as observed previously (23).

Together, these data verify that the *Mpa* NHEJ proteins possess similar biochemical activities to their bacterial counterparts *in vitro* and suggests that they perform equivalent roles in NHEJ-mediated DSB repair processes in this species, implying that this DSB repair pathway also operates in archaea. This NHEJ apparatus provides the first tractable *in trans* model system to study how

components of the archaeo-prokaryotic (AP) NHEJ apparatus co-ordinate the repair of DSBs *in vitro*.

Mpa* Ku is required for DNA end-joining *in vitro

Ku's proposed cellular function is to promote end-synapsis and recruit other NHEJ repair factors to the break. We therefore assayed the necessity for *Mpa* Ku to facilitate end-joining by *Mpa* Lig. The initial assay was performed on ds DNA containing a 3'-monoribonucleotide and a four base 5' DNA overhang with a self-annealing sequence to form a microhomology thus creating a doubly nicked duplex (Fig. 2A). *Mpa* Lig alone was not sufficient to ligate these breaks (Fig. 2A) however, incubation with Ku allowed successful ligation. The amount of product was proportional to the concentration of Ku added to the reactions. *Mpa* Lig's end-joining activity was also assayed using blunt-end ds DNA, containing a 3'-monoribonucleotide and a 5'-phosphate (Fig. 2B). Again, *Mpa* Lig alone was unable to ligate this substrate, however Ku stimulated its end-joining activity. Ligation of the blunt-ended DNA was not as efficient as that of the "synapsed" DNA, which is expected given the increased stability imparted by the presence of a region of microhomology.

Whilst it had been determined that *Mpa* Pol was not necessary for the physical act of connecting DNA ends to achieve ligation, a terminal ribonucleotide remained essential for ligation (Fig. S6). To ascertain whether *Mpa* Pol could efficiently fill a gap following break annealing, to promote ligation, we designed a DNA terminus with a 5 base 3' DNA overhang with a self-annealing sequence of four

bases (Fig. 2C). The non-complementary base requires gap-filling, in a template-dependent manner (S11A, S11C), subsequent to break annealing, before ligation can occur. These results indicate nucleotide incorporation by *Mpa* Pol and subsequent ligation of RNA/DNA nicks. The successful ligation of this substrate was again dependent on the presence of Ku for end-joining. Together, these data verify that *Mpa* Pol, Lig and Ku are *bona fide* NHEJ repair proteins and establish that a functional NHEJ complex exists in an archaeal species.

Phosphoesterase removal of 3'-phosphate assists DNA gap-filling and ligation

As discussed, *Mpa* PE possesses both phosphodiesterase and monophosphatase activities. PE is capable of removing a 3'-phosphate group from any DNA primer strand configuration (Fig. S7). This confirms that, unlike its ribonuclease activity, monophosphoesterase activity does not require a 2'-hydroxyl (OH) ribose moiety on the penultimate nucleotide. PE's ability to remove 3'-phosphates, in a context that is different to its ribonuclease activity, suggests that PE's phosphatase activity may participate in distinct processing events during end-joining. The likely role of the phosphatase is to facilitate 3' phosphate removal at damaged termini to expose a free 3'-OH, thus allowing the DNA termini to be receptive to subsequent modifications by the polymerase and ligase whose activities are dependent on the availability of a OH group.

To scrutinize this further, we incubated a single nucleotide gapped DNA substrate, containing both 3' and 5'-phosphate groups, with *Mpa* Pol, Lig and PE (+/-) (Fig. 2D). Double-stranded DNA with gaps are frequently used as substrates for

NHEJ enzymes as they mimic annealed breaks formed during DSB repair (13, 24). As predicted, the presence of a 3'-phosphate prevented NTP incorporation by *Mpa* Pol (Fig. 2D), due to the lack of a 3'-OH required for nucleophilic attack during extension. *Mpa* Lig was also unable to ligate this DNA, as ligation mechanisms also requires a free 3'-OH. This impediment was overcome by the addition of PE, which catalysed removal of the 3'-phosphate producing a phosphate shift (annotated as n-OH; Fig. 2D) followed by subsequent gap-filling and ligation. These data demonstrate that *Mpa* PE performs additional DNA modification tasks in NHEJ, aside from its ribonuclease function, thus acting as a phosphatase that can operate at early stages of DSB repair to facilitate further processing and ligation of the break termini.

***Mpa* Pol, a non-processive polymerase with strand displacement activity**

As discussed, *Mpa* Pol incorporated NTPs onto a primer strand in a template-dependent, but non-processive, manner (Fig. 1D) *Mpa* Pol inserted only 3-4 NTPs opposite a template strand (19 bases long) and efficiency was even further reduced with dNTPs, as noted above. This limited extension may result from the formation of DNA/RNA hybrids, switching the DNA conformation from B- to A-form, thus encouraging polymerase dissociation (25). *Mpa* Pol is also capable of catalyzing non-templated NTP incorporation on blunt-ended DNA, another activity shared with bacterial NHEJ polymerases (Fig. S8) (13).

To determine if *Mpa* Pol displays strand displacement activity following gap-filling (13, 26), we measured incorporation on a single nucleotide gap DNA substrate.

Although *Mpa* Pol efficiently filled in the gap, it also synthesised additional longer products, incorporating up to three nucleotides (Fig. 3A). This implies that, following gap-filling, *Mpa* Pol proceeds to incorporate further NTPs, presumably by displacing the downstream (D) strand. Notably, when *Mpa* Pol was assayed with a nicked DNA substrate (Fig. 3B), containing no gap, it also inserted up to 3 nucleotides. Even with nicked DNA, the NTP was incorporated in a template-dependent manner (Fig. S11B). Strand displacement synthesis is also intrinsic to the mycobacterial NHEJ polymerases (*Mtu* PolDom, Fig. S9) and has also been described in other DNA repair pathways (27). It is unclear what precise role displacement synthesis plays during NHEJ repair, at least in the absence of a 5' exonuclease, which could resect the displaced DNA flap. However, in addition to gap-filling, these polymerases also play important roles in end-synapsis, by directly facilitating the annealing of break termini (24). Therefore, Pol may maintain intrinsic displacement activity to enable it to open up breaks with limited complementarity, e.g. blunt ends. Pol may ingress into the termini, to uncover regions of microhomology that can then be annealed together to facilitate end-synapsis (24). It may also enable these enzymes to locate an internal 5' phosphate, if a terminal one is not present, to allow it to bind more securely to a DSB (24, 28). However, although it may be beneficial for some purposes, strand displacement can lead to some undesirable consequences, discussed below.

Phosphoesterase removes ribonucleotides from annealed DNA breaks

Although PE resects terminal ribonucleotides from recessed 3' primer strands (Fig. 1D), the specific role of this nucleolytic activity in processing NHEJ intermediates has

remained elusive. Given that polymerase incorporation of NTPs is the likely source for the presence of small tracts of RNA at sites of DSBs, we asked if PE's ribonuclease activity is required for processing these DNA-RNA intermediates? To address this question, we assayed PE's ability to resect an annealed DNA break substrate containing RNA, a diribonucleotide at the 3' end of the primer strand, and a D-strand. Notably, PE readily removed NMPs from the 3' DNA-RNA primer strand on this "annealed" break, leaving a single 3' ribonucleotide. Significantly, PE's resection activity was far more efficient than on an equivalent 3' primer lacking an adjacent D-strand (Fig. 3C, 3D).

Given the striking stimulation of PE's resection activity in the presence of a D-strand, we assayed a variety of different D-strands (ranging from a 2 base 5' flap (+2) to a 2 base (-2) gap) to identify PE's preferred substrate (Fig. 3E). A nicked substrate was the most efficiently processed intermediate, with almost 100% resection observed. A one base 5' flap (+1) and gapped substrate were processed less efficiently but considerably better than the remaining substrates (Fig. 3E). Notably, the activity of PE decreased drastically if the gap was larger than +1, whereas a 2 base 5' flap did not greatly suppress activity. These data demonstrate that PE is most efficient at resecting RNA incorporation at annealed breaks with little or no gap, even preferring to remove RNA that is displacing DNA rather than from a gap of 2, or more, bases.

Together, these data establish that PE preferentially resects RNA from the 3' primer strand when a D-strand is present, and not from a recessed 3' strand of a free

end, as previously reported (14, 29, 30). This observed preference for 3' resecting RNA from 'nicked' DNA supports a model in which PE specifically operates, not on break termini, but after DSB synapsis and gap filling has already occurred.

PE regulates incorporation of RNA at annealed breaks

The roles of NHEJ Pol and PE are antagonistic, as one inserts ribonucleotides at DSBs whilst the other one removes them. It is therefore critical to understand the context and co-ordination of these opposing end-processing activities to delineate how they co-operate to repair DSBs. As shown, PE prefers to resect RNA from the 3' end of primer strands at annealed breaks, indicating that it may be required to modulate the length of extension products incorporated during gap-filling. To test this model, a fixed concentration of PE was incubated with an increasing concentration of Pol and two different DNA substrates (+/- D-strand) containing a 1-nucleotide gap to assess the effects on NTP incorporation, if any (Fig. 4A, 4B). Notably, PE limited the amount of NTP incorporation, at approximately equimolar concentrations, on both substrates. PE significantly limited NTP incorporation in the presence of a D-strand, with the major product at +1, even at saturating Pol concentrations (Fig. 4B). These data establish that PE directly modulates the length of extension of primer strands at annealed DSBs by resecting the 3' RNA products incorporated by NHEJ polymerases.

PE resection of strand displacement intermediates during NHEJ repair

Although NHEJ polymerases are proficient at filling in gaps formed at annealed breaks (24), they also catalyse, less desirably, strand displacement synthesis on

gapped substrates. This is a particularly dangerous activity to possess given that it can lead to the formation of unligatable intermediates, containing a 5' flap, that are refractory to repair. A mechanism is therefore required to minimize the production of such abortive genotoxic intermediates, allowing end-ligation to proceed in a favourable way.

Given that PE regulates the amount of nucleotides incorporated by Pol into gapped NHEJ substrates, we assessed if PE provides a mechanism for modulating strand displacement synthesis thus preventing the formation of unligatable DNA intermediates (Fig. 4C, left panel). A single nucleotide gapped substrate was pre-incubated with *Mpa* Pol (0, 15, 30 and 60 minutes), with manganese and NTPs, before addition of *Mpa* PE and Lig. Whilst NTP incorporation by Pol remained at +3 or +4 across the time course, the amount of ligated product reduced dramatically over time. For example, after 15 minutes *Mpa* Pol had catalyzed almost total strand displacement synthesis, rendering most of the DNA inert to ligation (Fig. 4C, left panel). However, addition of *Mpa* PE (after 15, 30 or 60 minutes) rescued a considerable proportion of the non-repairable DNA (Fig. 4C, left panel), presenting *Mpa* Lig with a nicked substrate optimised for ligation. Indeed, addition of PE directly following addition of Pol (0 min; Fig. 4C) yielded a much greater quantity of ligated substrates than present in reactions lacking it. Similar results were also observed for *Mtu* PolDom (Fig. S10). End-joining substrates with 3' overhangs were also assayed as described above (S12A, S12B) and the results are consistent with the gap-filling and ligation assays in figure 4C, with PE counteracting Pol's downstream displacement activity. Together, these data robustly support a model in which PE is

required to optimally resect RNA-containing NHEJ intermediates, particularly those produced by strand displacement synthesis, to ensure the remodelled annealed break is amenable to ligation (Fig. 4C, right panel). The availability of this novel editing mechanism may be particularly important for ensuring the repair of all DSBs as even a single unrepaired DNA break can be lethal.

Concluding remarks

The potentially lethal nature of DNA double-strand breaks has ensured that all organisms have evolved repair pathways to effectively mend these potent lesions. Although recombination-based repair mechanisms are conserved across all living organisms, NHEJ break repair pathways were considered to operate exclusively in eukaryotes. However, in recent years it has been recognized that most bacterial species also utilize a closely related NHEJ repair apparatus to mend breaks. In contrast, there was limited evidence to support the existence of NHEJ repair pathways in the third major domain of life, archaea. This is particularly surprising given that many archaeal species thrive in some of the harshest environmental niches, e.g. deep sea thermal vents, significantly putting their genomes at increased risk of breakage. This current study resolves this apparent discrepancy with the discovery of a fully functional NHEJ repair apparatus in archaea thus establishing the conservation of this pathway across all three domains of life.

Although many archaeal DNA metabolism pathways, e.g. DNA replication, share many more common genes with eukaryotes rather than with bacteria, archaeal NHEJ is much more homologous to the bacterial DSB repair complexes, particularly

in terms of its conservation of the key repair proteins and their common intrinsic biochemical activities. Therefore, these closely related break repair pathways can be grouped together as archaeo-prokaryotic (AP) NHEJ to differentiate it from the more divergent eukaryotic pathway. Notably, a complete LigD AP-NHEJ apparatus is present in some plants, e.g. castor bean, suggesting that this pathway diverged quite late in evolution. The discovery of a shared end-joining repair apparatus in both archaeal and prokaryotic organisms further supports the hypothesis that a primordial NHEJ repair pathway arose early in evolution, was maintained and further evolved in lower eukaryotes, before eventually becoming the predominant DSB repair pathway in mammalian cells.

Although bacterial LigD has proven invaluable in providing insights into AP-NHEJ repair mechanisms, it also has some major experimental limitations. However, unlike LigD, the *Mpa* archaeon end-joining machinery is not fused into a single protein. This physical disconnectivity of the NHEJ activities has provided a tractable *in trans* repair apparatus that now allows dissection of the interplay between the end-processing enzymes to elucidate how they co-operate together to co-ordinate break repair. Although the AP-NHEJ phosphoesterase family was identified over a decade ago, its specific role in break repair has remained elusive. The modularity of the archaeal end-joining complex has enabled us to delineate a major function for PE in break repair. We demonstrate that the propensity of NHEJ polymerases to insert excessive ribonucleotides into gaps by strand displacement synthesis, following end-synapsis, is strongly counteracted by the resection activity of PE. NHEJ phosphoesterases limit excessive ribonucleotide incorporation by using a

ribonucleolytic resection mechanism, thus ensuring that nicked DNA substrates are available for ligation, preventing the accumulation of potentially genotoxic NHEJ intermediates (Fig. 5). The principle of strand displacement in DNA repair is not novel and, in fact, is well established in the eukaryotic base-excision repair (BER) pathway (27). Pol β actively displaces downstream damaged DNA as an interim repair step of BER, before FEN-1 endonucleolytically resects the 5' damaged flap. The PE activity associated with AP-NHEJ offers an alternative resolution of strand displaced intermediates. Resection of the newly synthesised ingressing DNA strand by PE facilitates reannealing of the displaced strand to produce a nicked template that is optimal for end-ligation.

Replicative DNA polymerases possess a 3' proofreading exonuclease (Exo) function to maintain fidelity. AP-NHEJ polymerases also maintain an associated 3'-Exo activity but to limit displacement synthesis rather than for maintaining fidelity. This functional association of these atypical nucleases with Pol and Lig is supported by the genomic distribution of *PE* genes, which are almost exclusively found in an operonic arrangement with *Pol* and *Lig* genes, indicating that a selective genetic requirement maintains PE in association with these NHEJ processing enzymes. PE-like genes also exist in some yeast species although their role is unknown.

An intriguing feature of AP-NHEJ repair pathways is the preferential insertion, removal and ligation of ribonucleotides at DSBs by the break repair enzymes. Utilisation of ribonucleotides during break repair is a concept of growing recognition with reports of preferentially incorporation of ribonucleotides during eukaryotic NHEJ

(31, 32), suggesting that RNA-DNA intermediates are not limited to AP-NHEJ (33). There are several reasons why it may be favourable to incorporate RNA during DSB end-joining. NHEJ is the DSB repair pathway of choice for non-dividing cells (10), where NTPs are much more abundant than dNTPs in the available nucleotide pools (33). Another distinct possibility, driven by the availability of ribonucleotides, is that NHEJ repair pathways evolved to selectively incorporate RNA, over DNA, into repaired DSBs to demarcate the boundaries of any alterations made to the repaired sites. This “labelling” of the repaired break with ribonucleotides may be useful in a number of ways. Firstly, It prevents the loss of genetic material at sites of DSBs by preventing resection of the DNA termini by the ribo-centric NHEJ repair machinery. Significantly in this regard, PE cannot resect DNA ends and prefers to resect short RNA tracts incorporated by the polymerase. Secondly, efficient ligation of breaks is also predicated on incorporation of a terminal 3' ribonucleotide that, potentially, acts as a molecular signal to indicate when a nick is ready for ligation thus preventing aberrant end-joining. A major consequence of AP-NHEJ repair is the incorporation of, at least, one ribonucleotide into ligated DSB junctions and it is likely that this RNA is subsequently excised from the repaired breaks (33, 34). Studies in murine systems have revealed that RNase HII performs an essential role in ribonucleotide removal from genomic DNA (35). Whilst this role is thought largely to protect from RNA misincorporated during replication, RNase HII (or a similar enzyme) may also resolve the final complication of AP-NHEJ repair but this remains to be established.

Methods

Purification of *Mpa* Lig, Pol, PE and Ku proteins

All *Mpa* NHEJ genes were cloned into pET 28A, the histidine-tagged proteins expressed in B834S (DE3) and purified using the same chromatography protocols, except where noted. Pelleted cells containing the appropriate plasmid were grown in Terrific Broth and pelleted cultures resuspended in lysis buffer (50mM Tris-HCl pH7.5, 500mM NaCl, 30mM Imidazole, 10% glycerol, 17mg/mL PMSF, 34mg/mL benzamidine). Cells were disrupted by sonication and soluble proteins were isolated by centrifugation. The supernatant was applied to a 25mL NTA (Qiagen) column equilibrated in IMAC buffer A (as lysis buffer) and eluted in IMAC buffer B (as A, with 300mM imidazole). *Mpa* Ku supernatant was subjected to an extended wash period in IMAC buffer C (as A, with 1M NaCl) to remove excess DNA bound to Ku proteins. Protein containing fractions were loaded onto 5mL fast flow Q and S (Lig, Pol), or Q (PE and Ku) columns (GE Healthcare) pre-equilibrated in IEx buffer A (50mM Tris-HCl pH 7.5, 10% glycerol) for separation by ionic charge. Lig, Pol and PE were eluted in the flow through without binding the column, Ku was eluted in 10% IEx buffer B (50mM Tris-HCl pH7.5, 2M NaCl, 10% glycerol). Removal of any remaining contaminating proteins was achieved by gel filtration on a Superdex S200 column (GE Healthcare) equilibrated in GF buffer (25mM Tris pH 7.5, 500mM NaCl, 10% glycerol). Eluted proteins were analysed for purity using SDS-PAGE and concentrations ascertained by spectrophotometry of samples at 280nm.

EMSA

Fluorescein labelled DNA primer (5'-CATATCCGTGTCGCCCCTTATTCCGATAGTGACTACA) was annealed to DNA template (5'-TGTAGTCACTATCGGAATAAGGGGCGACACGGATATG) creating Oligo 20 for DNA binding assays. 60nM DNA was incubated with 0.2-1.6 μ M *Mpa* Ku in 50mM Tris pH 7.5 and 5% glycerol for 30 minutes at room temperature in a volume of 20 μ L. Samples were separated by electrophoresis on a native 5%

polyacrylamide gel in 0.5x TBE buffer for 2hr. Fluorescently labelled DNA oligomers were detected by scanning using a Fujifilm FLA-5100.

DNA ligation assay

Fluorescein labelled DNA primer (5'- CTATGAGCGAATCGCrC) was annealed to DNA template (5'-AGTCGCATAGTGTAGTCGGGGCGATTTCGCTCATAG) with a downstream sequence (5'-P-CGACTACACTATGCGACT) creating Oligo 60. 30nM DNA was incubated with 300nM *Mpa* Lig in 50mM Tris HCl pH 7.5, 5mM MnCl₂ for 30, 60 and 90 minutes at 37°C in a volume of 20μL. The reactions were stopped by addition of Stop buffer (95% formamide, 0.09% xylene cynaol) and boiled at 95°C for 10 minutes. Samples were separated by electrophoresis on an 8M Urea, 15% polyacrylamide gel in 1xTBE buffer for 2hr. Fluorescently labelled oligos were detected by scanning using a Fujifilm FLA-5100.

DNA extension assay

Fluorescein labelled DNA primer (5'-CTATGAGCGAATCGCC) was annealed to DNA template (5'-AGTCGCATAGTGTAGTCGGGGCGATTTCGCTCATAG) creating Oligo 9 for extension. To this DNA oligo, further downstream sequences were annealed (5'-CCGACTACACTATGCGAC, 5'-P-CGACTACACTATGCGACT) creating Oligos 11 and 37 for gap filling assays. 30nM DNA was incubated with 300nM *Mpa* Pol in 50mM Tris HCl pH 7.5, 5mM MnCl₂, 250μM NTP mix for 1hr at 37°C in a volume of 20μL. The reactions were stopped by addition of Stop buffer (95% formamide, 0.09% xylene cynaol) and boiled at 95°C for 10 minutes. Samples were separated by electrophoresis on an 8M Urea, 15% polyacrylamide gel in 1xTBE buffer for 2hr. Fluorescently labelled oligos were detected by scanning using a Fujifilm FLA-5100.

Ribonuclease and phosphatase assay

Fluorescein labelled DNA-RNA primer (5'-CTATGAGCGAATCGrCrC) was annealed to DNA template (5'-AGTCGCATAGTGTAGTCGGGGCGATTTCGCTCATAG) creating Oligo 3 for phosphoesterase assays. To this oligo further downstream sequences were annealed (5'-CTACCGACTACACTATGCGACT 5'-CCGACTACACTATGCGACT 5'-P-CGACTACACTATGCGACT 5'- GACTACACTATGCGACT) creating oligos 25, 62, 2, 38 and 43 respectively. 30nM DNA was incubated with 300nM *Mpa* PE in 50mM Tris pH 7.5 and 5mM $MnCl_2$ at 37°C for varying time points across 90 minutes. The reactions were quenched by addition of Stop buffer (95% formamide, 0.09% xylene cynaol) and boiled at 95°C for 10 minutes. Samples were separated by electrophoresis on an 8M Urea, 15% polyacrylamide gel in 1xTBE buffer for 2hr. Fluorescently labelled oligos were detected by scanning using a Fujifilm FLA-5100.

DNA end joining assay

Fluorescein labelled DNA and DNA-RNA primers (5'-CATATCCGTGTCGCCCCTTATTCCGATAGTGACTArC 5'-CATATCCGTGTCGCCCCTTATTCCGATAGTGACTACrA 5'-CATATCCGTGTCGCCCCTTATTCCGATAGTGACTACAACGCG) were annealed to DNA templates (5'-P-CATGGTAGTCACTATCGGAATAAGGGGCGACACGGATAT 5'-P- TGTAGTCACTATCGGAATAAGGGGCGACACGGATATG) to form Oligos 45, 68 and 76 respectively, for DNA end joining assays. 300nM *Mpa* Lig was incubated with 200-800nM *Mpa* Ku, and 300nM *Mpa* Pol where indicated, in 50mM Tris pH 7.5, 5mM $MnCl_2$, 1mM DTT, 5% glycerol, 0.1mgml⁻¹ BSA, 250μM NTPs (Pol assay only) at 37°C in a final volume of 20μL. The reactions were quenched by addition of Stop buffer (95% formamide, 0.09% xylene cynaol) and boiled at 95°C for 10 minutes. Samples were separated by electrophoresis on an 8M Urea, 15% polyacrylamide gel in 1xTBE buffer for 2hr. Fluorescently labelled oligos were detected by scanning using a Fujifilm FLA-5100.

Phosphatase, gap filling and ligation assay

Fluorescein labelled DNA primer (5'-CTATGAGCGAATCGCC-P) was annealed to DNA template (5'-AGTCGCATAGTGTAGTCGGGGCGATTTCGCTCATAG) and downstream strand (5'-P-CGACTACACTATGCGACT) creating Oligo 63 for phosphatase, gap filling and ligation. 300nM of *Mpa* Lig, Pol and PE were incubated where indicated with 50mM Tris HCl pH 7.5, 5mM MnCl₂, 250μM NTP mix for 1hr at 37°C in a volume of 20μL. The reactions were quenched by addition of Stop buffer buffer (95% formamide, 0.09% xylene cynaol) and boiled at 95°C for 10 minutes. Samples were separated by electrophoresis on an 8M Urea, 15% polyacrylamide gel in 1xTBE buffer for 2hr. Fluorescently labelled oligos were detected by scanning using a Fujifilm FLA-5100.

Polymerase DNA extension, gap-filling and phosphoesterase assay

Oligos 9, 11 and 37, as described above, were used for DNA extention and phosphoesterase assays. 30nM DNA was incubated with 300nM *Mpa* Pol and PE where indicated in 50mM Tris pH 7.5 and 5mM MnCl₂ at 37°C for varying time points across 90 minutes. The reactions were quenched by addition of Stop buffer buffer (95% formamide, 0.09% xylene cynaol) and boiled at 95°C for 10 minutes. Samples were separated by electrophoresis on an 8M Urea, 15% polyacrylamide gel in 1xTBE buffer for 2hr. Fluorescently labelled oligos were detected by scanning using a Fujifilm FLA-5100.

DNA gap-filling, phosphoesterase and ligation assay

Oligo 37, as described above, was used for gap-filling, phosphoesterase and ligation assays. 30nM DNA was incubated with 300nM *Mpa* Pol in 50mM Tris pH 7.5 and 5mM MnCl₂ at 37°C for 0, 15, 30 and 60 minutes. 300nM *Mpa* Lig and PE were then added and incubation continued for a further 60 minutes. The reactions were

quenched by addition of Stop buffer buffer (95% formamide, 0.09% xylene cynaol) and boiled at 95°C for 10 minutes. Samples were separated by electrophoresis on an 8M Urea, 15% polyacrylamide gel in 1xTBE buffer for 2hr. Fluorescently labelled oligos were detected by scanning using a Fujifilm FLA-5100.

Acknowledgements

We declare that none of the authors have a financial interest related to this work. Work in the AJD laboratory is supported by grants from Biotechnology and Biological Sciences Research Council (BBSRC) and indirectly by a centre grant from the MRC. EB was supported by a BBSRC PhD studentship. We would like to thank Dr H. Imachi for providing *Methanocella paludicola* genomic DNA.

Conflict of interest

All authors declare no conflicts of interest.

References

1. Daley JM, Palmboos PL, Wu D, Wilson TE (2005) Nonhomologous end joining in yeast. *Annu Rev Genet* 39:431–451.
2. Helleday T, Lo J, van Gent DC, Engelward BP (2007) DNA double-strand break repair: from mechanistic understanding to cancer treatment. *DNA Repair* 6:923–935.
3. Mahaney BL, Meek K, Lees-Miller SP (2009) Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *Biochem J* 417:639–650.
4. Nick McElhinny SA et al. (2005) A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining. *Mol Cell* 19:357–366.
5. Weinfeld M, Mani RS, Abdou I, Aceytuno RD, Glover JNM (2011) Tidying up loose ends: the role of polynucleotide kinase/phosphatase in DNA strand break repair. *Trends Biochem Sci* 36:262–271.
6. Gu J et al. (2010) DNA-PKcs regulates a single-stranded DNA endonuclease activity of Artemis. *DNA Repair* 9:429–437.
7. Wu X, Wilson TE, Lieber MR (1999) A role for FEN-1 in nonhomologous DNA end joining: the order of strand annealing and nucleolytic processing events. *Proc Natl Acad Sci USA* 96:1303–1308.
8. Della M et al. (2004) Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science* 306:683–685.
9. Weller GR et al. (2002) Identification of a DNA nonhomologous end-joining complex in bacteria. *Science* 297:1686–1689.
10. Pitcher RS et al. (2007) NHEJ protects mycobacteria in stationary phase against the harmful effects of desiccation. *DNA Repair* 6:1271–1276.
11. Gong C et al. (2005) Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. *Nat Struct Mol Biol* 12:304–312.
12. Zhu H, Shuman S (2008) Bacterial nonhomologous end joining ligases preferentially seal breaks with a 3'-OH monoribonucleotide. *J Biol Chem* 283:8331–8339.
13. Pitcher RS et al. (2007) Structure and function of a mycobacterial NHEJ DNA repair polymerase. *J Mol Biol* 366:391–405.
14. Zhu H, Shuman S (2005) Novel 3'-Ribonuclease and 3'-Phosphatase Activities of the Bacterial Non-homologous End-joining Protein, DNA Ligase D. *Journal of Biological Chemistry* 280:25973–25981.
15. Aniukwu J, Glickman MS, Shuman S (2008) The pathways and outcomes of mycobacterial NHEJ depend on the structure of the broken DNA ends. *Genes Dev* 22:512–527.
16. Aravind L, Koonin EV (2001) Prokaryotic homologs of the eukaryotic DNA-end-binding

- protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res* 11:1365–1374.
17. Nair PA, Smith P, Shuman S (2010) Structure of bacterial LigD 3'-phosphoesterase unveils a DNA repair superfamily. *PNAS* 107:12822–12827.
 18. Smith P, Nair PA, Das U, Zhu H, Shuman S (2011) Structures and activities of archaeal members of the LigD 3'-phosphoesterase DNA repair enzyme superfamily. *Nucleic Acids Research* 39:3310–3320.
 19. Sakai S et al. (2008) *Methanocella paludicola* gen. nov., sp. nov., a methane-producing archaeon, the first isolate of the lineage “Rice Cluster I,” and proposal of the new archaeal order Methanocellales ord. nov. *Int J Syst Evol Microbiol* 58:929–936.
 20. Pitcher RS, Tonkin LM, Green AJ, Doherty AJ (2005) Domain structure of a NHEJ DNA repair ligase from *Mycobacterium tuberculosis*. *J Mol Biol* 351:531–544.
 21. Shuman S, Liu Y, Schwer B (1994) Covalent catalysis in nucleotidyl transfer reactions: essential motifs in *Saccharomyces cerevisiae* RNA capping enzyme are conserved in *Schizosaccharomyces pombe* and viral capping enzymes and among polynucleotide ligases. *Proc Natl Acad Sci USA* 91:12046–12050.
 22. Doherty AJ, Suh SW (2000) Structural and mechanistic conservation in DNA ligases. *Nucleic Acids Research* 28:4051–4058.
 23. Doherty AJ, Ashford SR, Subramanya HS, Wigley DB (1996) Bacteriophage T7 DNA ligase. Overexpression, purification, crystallization, and characterization. *J Biol Chem* 271:11083–11089.
 24. Brissett NC et al. (2007) Structure of a NHEJ polymerase-mediated DNA synaptic complex. *Science* 318:456–459.
 25. Zhu H, Bhattarai H, Yan H-G, Shuman S, Glickman MS (2012) Characterization of *Mycobacterium smegmatis* PolD2 and PolD1 as RNA/DNA polymerases homologous to the POL domain of bacterial DNA ligase D. *Biochemistry*.
 26. Zhu H, Shuman S (2010) Gap filling activities of *Pseudomonas* DNA ligase D (LigD) polymerase and functional interactions of LigD with the DNA end-binding Ku protein. *Journal of Biological Chemistry* 285:4815–4825.
 27. Imoto S, Bransfield LA, Croteau DL, Van Houten B, Greenberg MM (2008) DNA tandem lesion repair by strand displacement synthesis and nucleotide excision repair. *Biochemistry* 47:4306–4316.
 28. Brissett NC et al. (2011) Structure of a preternary complex involving a prokaryotic NHEJ DNA polymerase. *Mol Cell* 41:221–231.
 29. Zhu H, Wang LK, Shuman S (2005) Essential constituents of the 3'-phosphoesterase domain of bacterial DNA ligase D, a nonhomologous end-joining enzyme. *J Biol Chem* 280:33707–33715.
 30. Zhu H, Shuman S (2006) Substrate specificity and structure-function analysis of the 3'-phosphoesterase component of the bacterial NHEJ protein, DNA ligase D. *J Biol*

Chem 281:13873–13881.

31. Martin MJ, Garcia-Ortiz MV, Esteban V, Blanco L (2012) Ribonucleotides and manganese ions improve non-homologous end joining by human Polμ. *Nucleic Acids Research*.
32. Nick McElhinny SA, Ramsden DA (2003) Polymerase mu is a DNA-directed DNA/RNA polymerase. *Mol Cell Biol* 23:2309–2315.
33. Pitcher RS, Wilson TE, Doherty AJ (2005) New insights into NHEJ repair processes in prokaryotes. *Cell Cycle* 4:675–678.
34. Rydberg B, Game J (2002) Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts. *Proc Natl Acad Sci USA* 99:16654–16659.
35. Reijns, MA et al. (2012) Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell* 149: 1008–1022.

Figure legends

Figure 1. Biochemical activities associated with *Mpa* NHEJ repair proteins. (A)

Comparison of gene arrangement in a NHEJ operons from *Methanocella paludicola* (*Mpa*) and *Mycobacterium tuberculosis* (*Mtu*). *Mpa* NHEJ factors are expressed as discrete proteins, whilst *Mtu* LigD is a multi-domain single polypeptide. Ku is present in both genomes, and is operonic with LigD in *Mtu*. (B) Ku DNA binding assays, 100nM 5'-fluorescein labelled 33mer dsDNA was incubated with 0, 200, 400, 800 and 1600nM *Mpa* Ku protein. (C) Ligase reactions contained 30nM 5'-fluorescein labelled 16mer DNA-RNA (D15R1) annealed to template DNA and 5'-phosphorylated D-strand to create a nicked substrate, and 300nM *Mpa* Lig protein. Ligated product is 35 bases in length. (D) Polymerase primer extension reactions contained 30nM 5'-fluorescein labelled 16mer DNA annealed to template DNA, and 300nM *Mpa* Pol protein. Incubations included 250 μ M NTPs/dNTPs and 5mM MnCl₂ or MgCl₂ where indicated. (E) Phosphoesterase reactions contained 30nM 5'-fluorescein labelled 16mer DNA/RNA (D14R2) annealed to template DNA, and 300nM *Mpa* PE protein. n-R-R is unmodified substrate, n-R-P represents removals of the terminal RMP to yield a terminal phosphate group, and n-R-OH is the product following phosphate removal leaving a terminal hydroxyl group. All reactions performed in the presence of manganese (5mM).

Figure 2. *Mpa* NHEJ complexes repair a diverse range of DNA breaks *in vitro*. (A)

End-joining reaction mixtures contained 30nM 5'-fluorescein labelled 36mer DNA-RNA (D35R1) annealed to 5'-phosphorylated 40mer template DNA to create self-complementary 4mer overhang, with 300nM *Mpa* Lig and 200-800nM *Mpa* Ku proteins. Ligated products were 76 bases in length. (B) End-joining reaction mixtures contained 30nM 5'-fluorescein labelled 36mer DNA-RNA (D35R1) annealed to 5'-phosphorylated 36mer template DNA to create blunt ended 5'-phosphorylated DNA, with 300nM *Mpa* Lig and 200-800nM *Mpa* Ku proteins. Ligated products were 72 bases in length. (C) End-joining reaction mixtures contained 30nM 5'-fluorescein labelled 42mer DNA annealed with 5'-phosphorylated 37mer template DNA to create 5mer 3'-overhang DNA, four bases of which are self-complementary, with 300nM *Mpa* Lig and Pol and 200-800nM *Mpa* Ku proteins. Ligated products were 79 bases

in length. **(D)** 3'-phosphatase, gap filling and ligation reaction mixtures contained 30nM 5'-fluorescein labelled 16mer with a 3'-phosphate annealed to template DNA and 5'-phosphorylated D-strand, with 300nM *Mpa* Lig, Pol and PE where indicated. n-P indicates unmodified substrate, n-OH indicates removal of 3'-phosphate to reveal 3'-OH which can then receive incoming NTPs for n+1. All reactions performed in the presence of manganese (5mM).

Figure 3. Characterization of polymerase strand displacement and phosphoesterase activities. **(A, B)** Gap filling reactions mixtures contained 30nM 5'-fluorescein labelled 16mer DNA annealed to template DNA and D-strands to create 1 nucleotide gap DNA **(A)** or nicked DNA respectively **(B)**, and 300nM *Mpa* Pol protein. **(C, D, E)** Phosphoesterase reaction mixtures contained 5'-fluorescein labelled DNA-RNA (D14R2) annealed to a template with no D-strand **(C)**, with D-strand **(D)**, with D-strands producing 2 and 1 base nucleotide 5' flaps and 1 and 2 nucleotide gaps in respect to the DNA-RNA primer **(E)** and 300nM *Mpa* PE. Products as described in Fig. 1E. All reactions performed in the presence of manganese (5mM).

Figure 4. PE resection of strand displacement intermediates is required for NHEJ repair. **(A, B)** Polymerase and phosphoesterase reaction mixtures contained 5'-fluorescein labelled 16mer DNA primers annealed to a template with no D-strand **(A)** and with a 1 nucleotide gap **(B)**, and 300nM *Mpa* PE and 0.1-1.4 μ M *Mpa* Pol. **(C, left panel)** Gap filling, phosphoesterase and ligation reaction mixtures contained 5'-fluorescein labelled 16mer DNA annealed to template DNA and a D-strand producing a single nucleotide gap, 300nM *Mpa* Lig, Pol and PE where indicated. Pre-incubated with *Mpa* Pol at 37°C for 0, 15, 30 and 60 minutes before addition of *Mpa* Lig and PE and incubation at 37°C for 1hr. All reactions performed in the presence of manganese (5mM). **(C, right panel)** Illustration of rescue of abortive NHEJ repair intermediates which cannot be ligated. Downstream DNA displaced by incorporation of NTPs prevents ligation, however *Mpa* PE resects the RNA tract and leaves a monoribonucleotide. Downstream DNA can then reanneal and the substrate is ready for ligation with the 5'-phosphate opposite the 3'-hydroxyl group.

Figure 5. Model of AP-NHEJ using RNA repair intermediates. The initial step in AP-NHEJ repair is the binding of Ku homodimers to broken ends and bringing them into close proximity. Ku and Pol promote microhomology-mediated synapsis of overhanging termini, if present, to anneal the break. Pol fills in any resulting gaps, at either end of the annealed break, by a template-dependent RNA synthesis. Pol may displace several bases of downstream DNA using its distinct ability to mobilise and direct DNA strands, offering a greater flexibility to repair a variety of configurations of damaged DNA. Strand displacement RNA synthesis by Pol is regulated by PE, which can remove unnecessary NMPs and allow the displaced DNA to realign with the template. PE always leaves a single terminal 3' ribonucleotide in place in resected breaks, presenting Lig with an optimised nicked substrate for ligation. Although the broken DNA is reconnected, repair may not be fully completed as several RNA bases still resided in the duplex. RNA removal may be performed by additional enzymes, such as RNase HII, before other repair polymerases fill in the gaps with DNA.

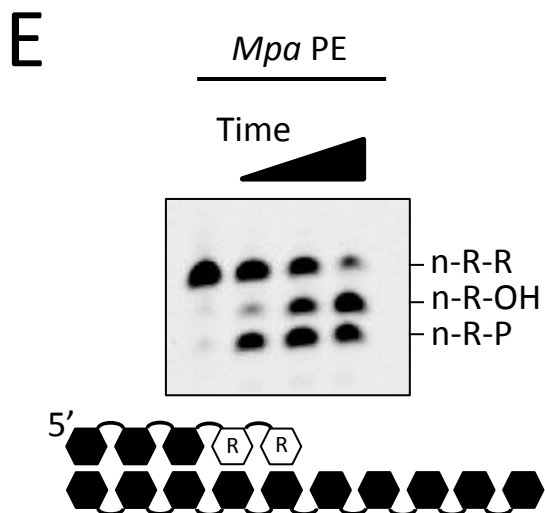
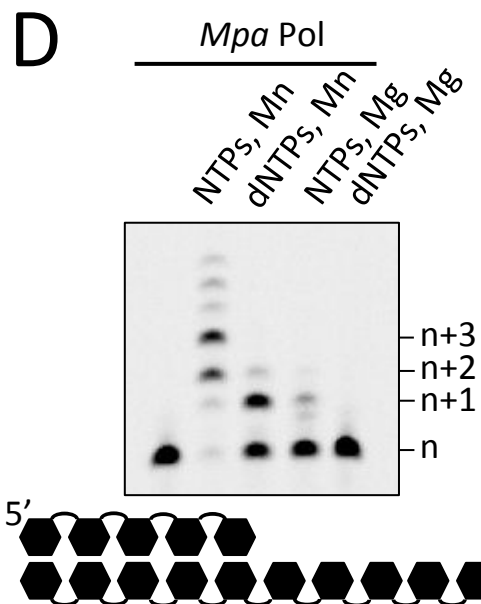
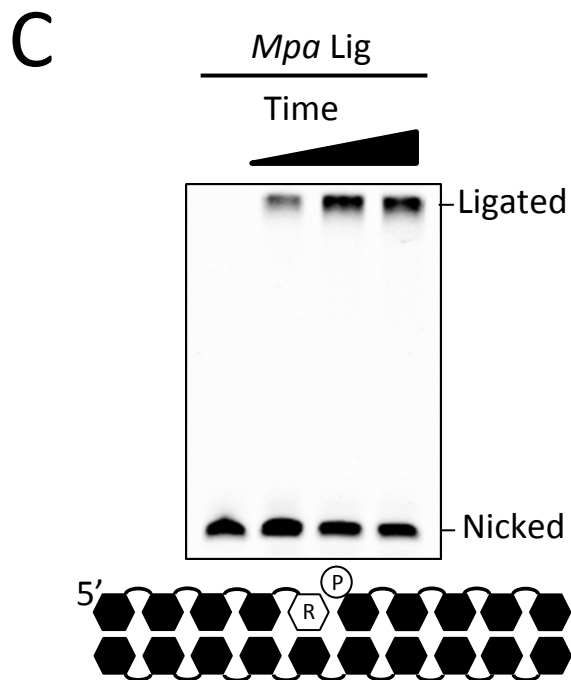
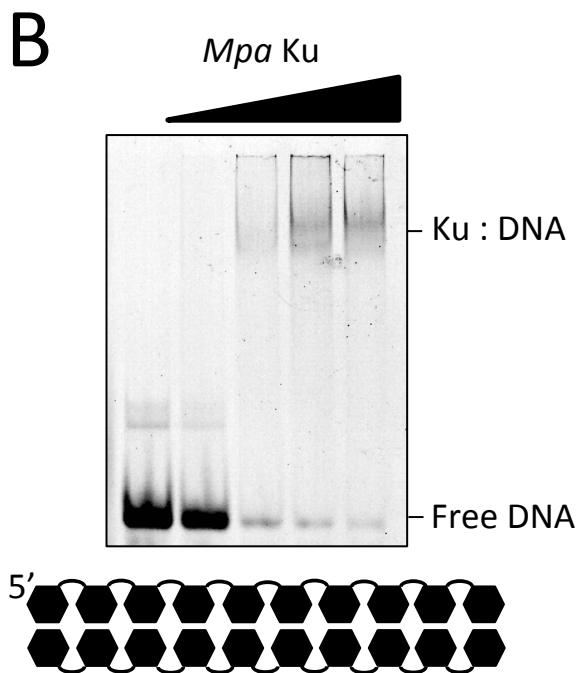
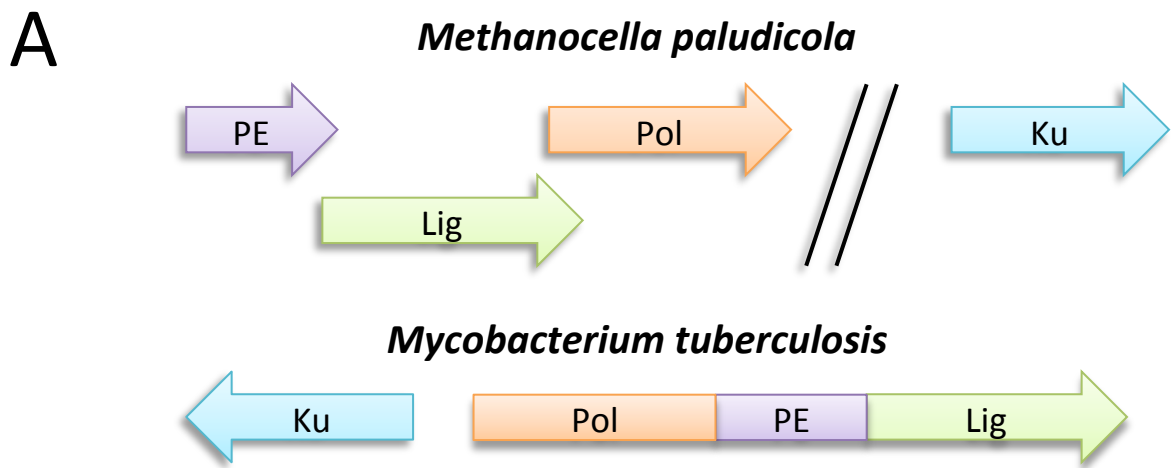
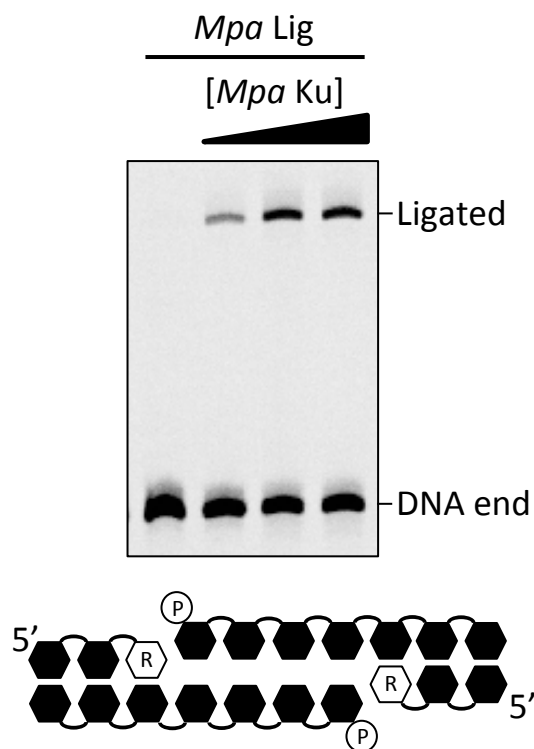
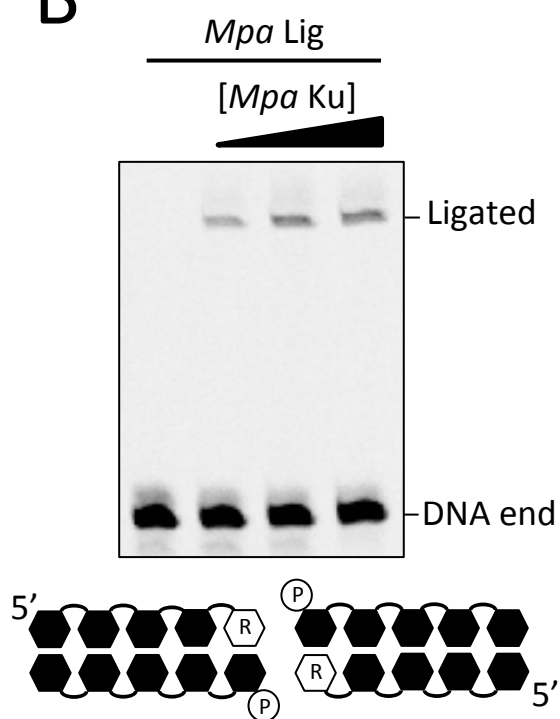


Figure 1

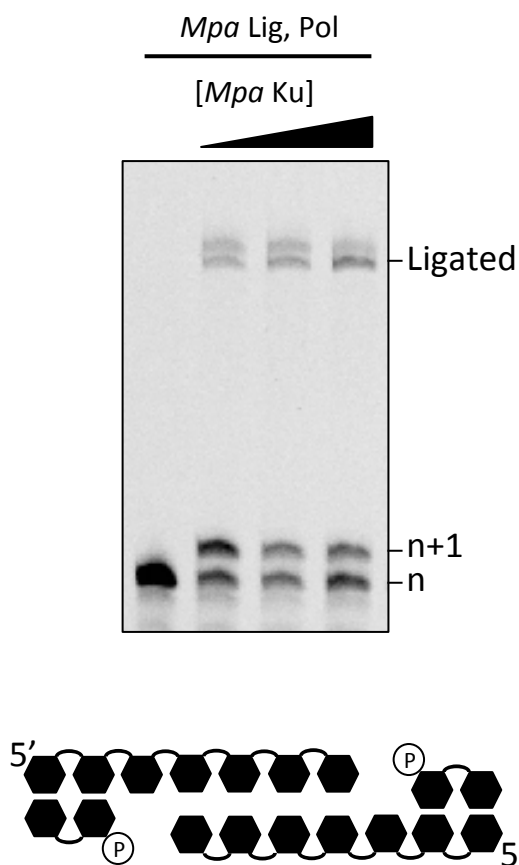
A



B



C



D

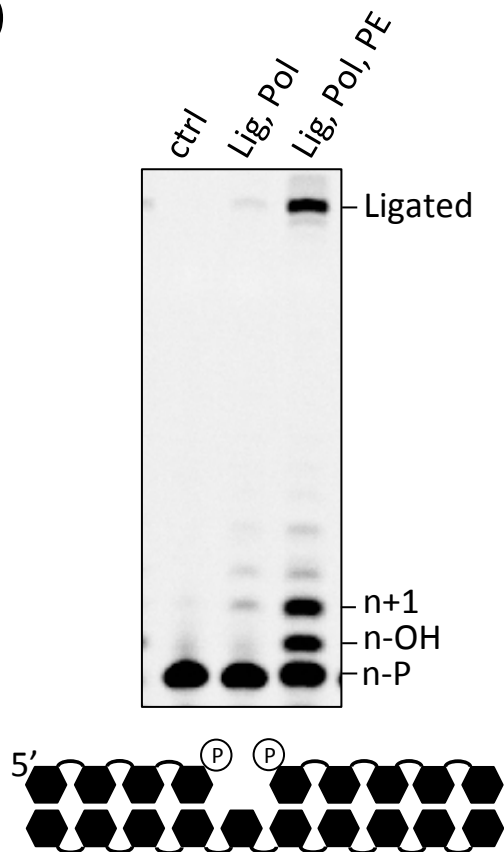


Figure 2

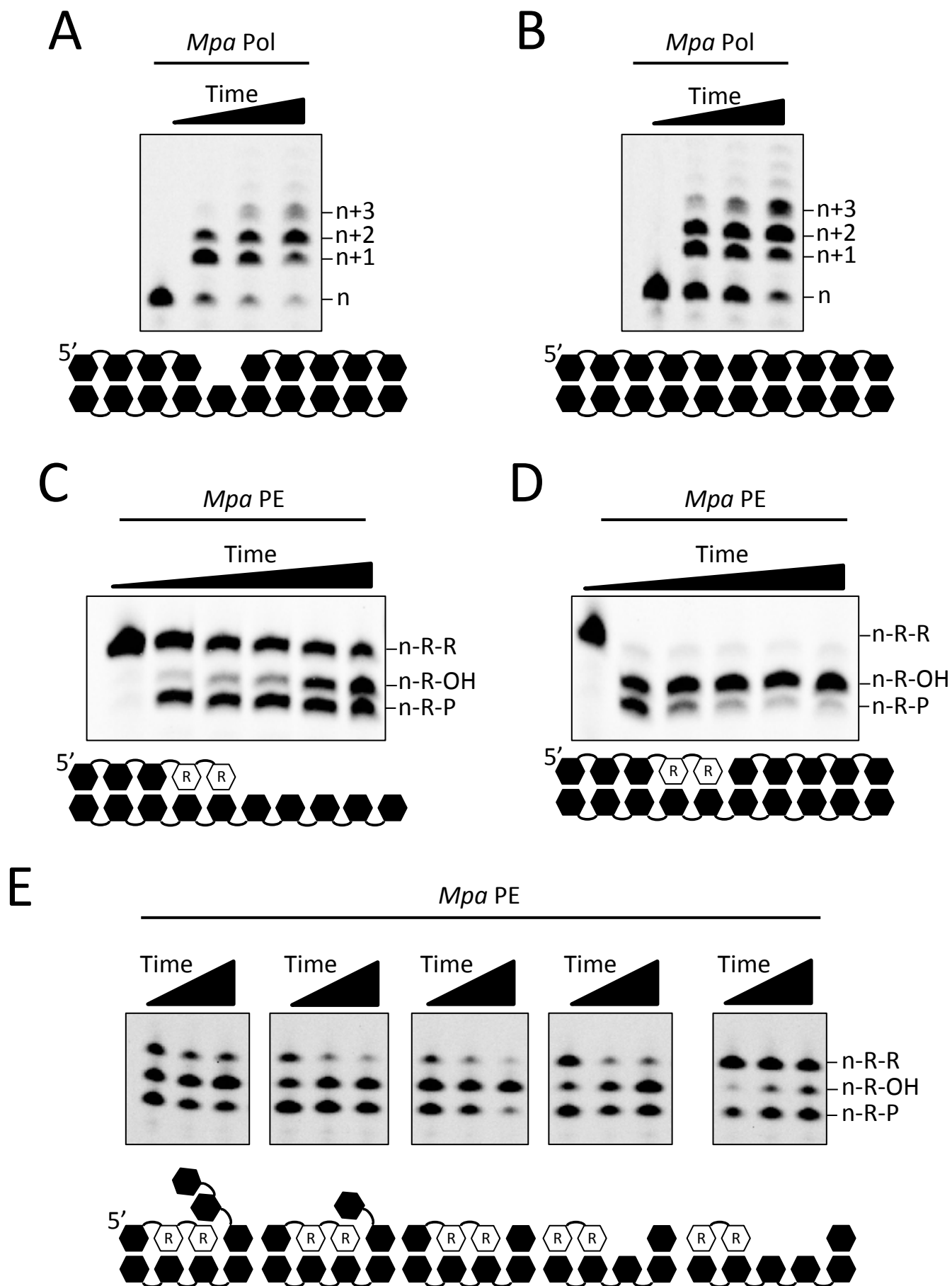


Figure 3

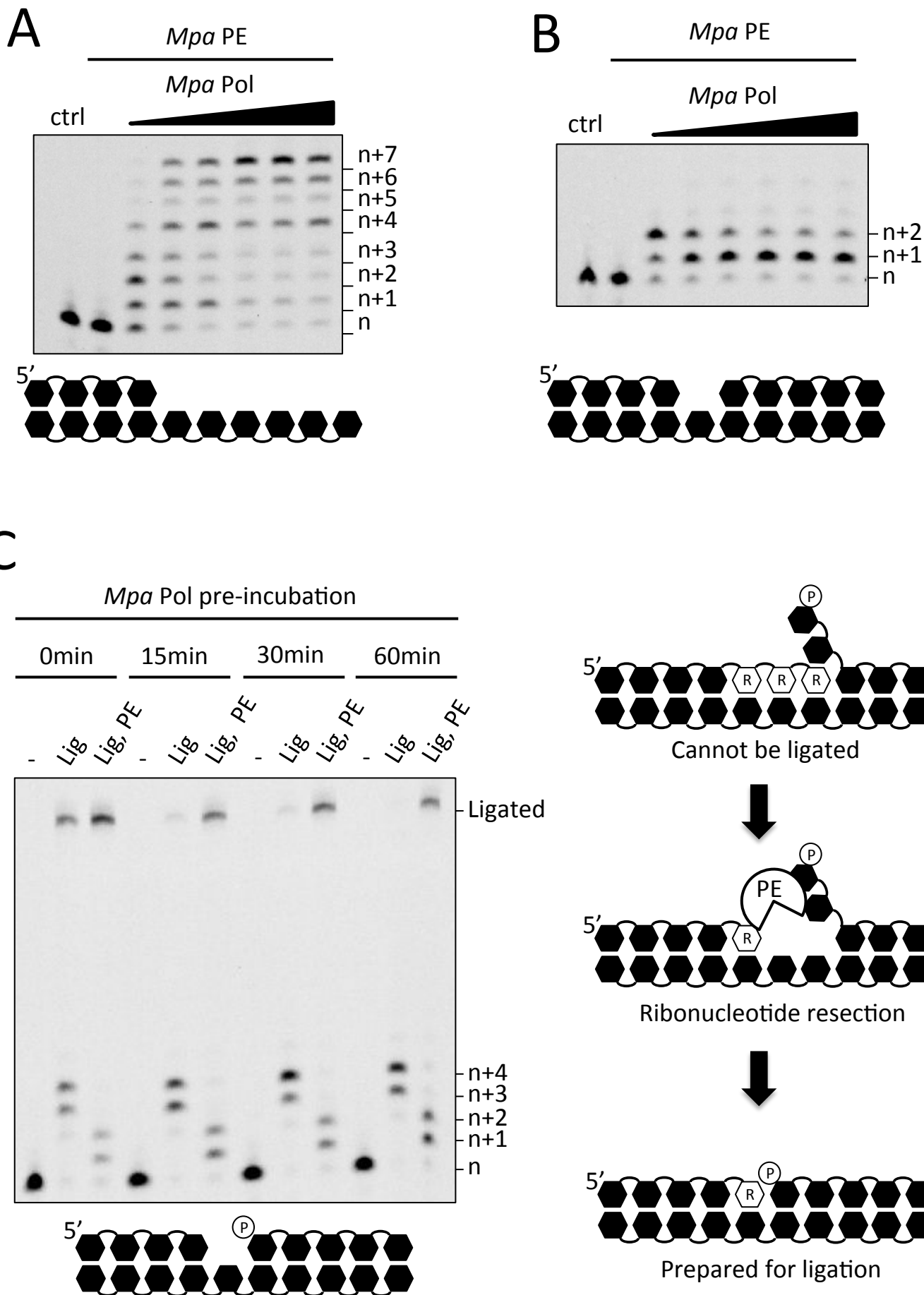


Figure 4

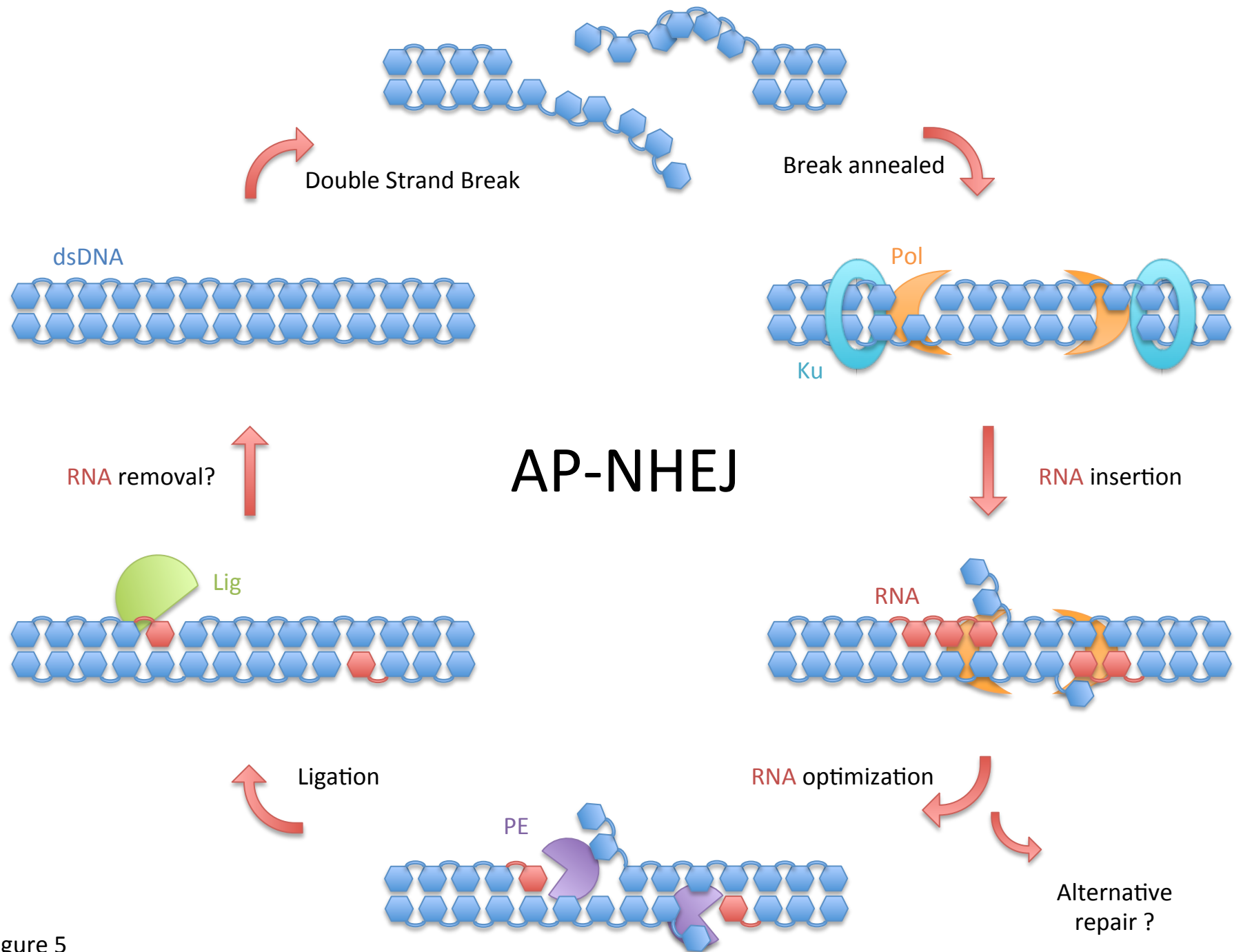


Figure 5

Supporting information:

Ribonucleolytic resection is required for repair of strand displaced NHEJ intermediates, Edward Bartlett, Nigel C. Brissett and Aidan J. Doherty

S1. A, B, C & D Protein sequence alignments of *Mpa* Lig, Pol, PE and Ku with *Mtu*, *Msm*, *Pae* LigD and Ku.

(A) *Methanocella paludicola* (*Mpa*) Ligase protein sequence alignment with Ligase D (LigD) ligase domains (LigDom) of *Mycobacterium tuberculosis* (*Mtu*), *Mycobacterium Smegmatis* (*Msm*) and *Pseudomonas Aeruginosa* (*Pae*). B. *Mpa* polymerase (Pol) protein sequence alignment with LigD polymerase domains (PolDom) of *Mtu*, *Msm* and *Pae*. C. *Mpa* phosphoesterase (PE) protein sequence alignment with LigD phosphoesterase domains (PEDom) of *Mtu*, *Msm* and *Pae*. D. *Mpa* Ku protein sequence alignment with Ku of *Mtu*, *Msm* and *Pae*. All alignments were made with TCoffee and edited in JalView.

S2. SDS-PAGE *Mpa* Lig, Pol, PE and Ku.

SDS-PAGE of gel filtration chromatography fractions of respective purifications of *Mpa* Lig, Pol, PE and Ku. Migration sizes of the proteins were: *Mpa* Lig ~40kD, *Mpa* Pol ~34kD, *Mpa* PE ~24kD and *Mpa* Ku ~30kD (monomer) ~60kD (dimer). The gels indicate the high purity of the proteins.

S3. Archaeo-eukaryotic primase (AEP) superfamily motif sequence alignment.

Three key conserved motifs (I, II and III) confer membership to the AEP superfamily. Here *Mpa* Pol is aligned with *Mycobacterium tuberculosis* and *Pseudomonas Aeruginosa* NHEJ Pols, and *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* primase small sub-units.

S4. *Mpa* PE ion preference.

Phosphoesterase reaction mixtures contained 50mM Tris pH 7.5, 5mM MnCl₂, MgCl₂, CoCl₂, CuCl₂, CaCl₂, NiCl₂ or Zn(O₂CCH₃)₂ where indicated, 30nM 5'-fluorescein labelled 16mer DNA/RNA (D14R2), and 300nM *Mpa* PE protein. Incubated at 37°C for 1 hour 30 minutes and electrophoresed on a 15% denaturing polyacrylamide gel.

S5. Polynucleotide ligase and mRNA capping superfamily motifs

Six key conserved motifs (I, Ia, III, IIIa, IV and V) confer membership to the polynucleotide ligase and mRNA capping superfamily. Motif VI is also included with highly conserved residues. Here *Mpa* Lig is aligned

Mycobacterium tuberculosis and *Pseudomonas aeruginosa* Ligase D; Enterobacterio phage T7, *Homo sapiens*, *Mus musculus* DNA ligase; and *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* mRNA capping enzymes.

S6. *Mpa* Lig preferable ligation with Mn, not Mg

Ligase reaction mixtures contained 50mM Tris pH 7.5, 5mM MnCl₂ or Mg Cl₂, 30nM 5'-fluorescein labelled 16mer oligomer (D16, D15R1,D14R2, D12R4 or D8R8 where indicated) and 19mer 5'-phosphorylated DNA, and 300nM *Mpa* Lig protein. Incubated at 37°C for 2 hours and electrophoresed on a 15% denaturing polyacrylamide gel.

S7. *Mpa* PE phosphomonoesterase activity does not require 2'-OH

Phosphatase reaction mixtures contained 50mM Tris pH 7.5, 5mM MnCl₂, 30nM 5'-fluorescein labelled 16mer DNA with 3'-phosphate, and 250, 500 and 1000nM *Mpa* PE protein. Incubated at 37°C for 1 hour 30 minutes and electrophoresed on a 15% denaturing polyacrylamide gel.

S8. *Mpa* Pol non-templated NTP incorporation

Polymerase non-templated extension reaction mixtures contained 50mM Tris pH7.5, 5mM MnCl₂, 62.5μM ATP, CTP, GTP, UTP or 250μM mixed NTPs where indicated, 30nM 5'-fluorescein labelled 36mer DNA, and 300nM *Mpa* Pol protein. Incubated at 37°C for 1 hour and electrophoresed on a 15% denaturing polyacrylamide gel.

S9. *Mtu* PolDom strand displacement

Polymerase strand displacement reaction mixtures contained 50mM Tris pH 7.5, 5mM MnCl₂, 250μM NTPs, 30nM 5'-fluorescein labelled 16mer DNA, and 300nM *Mpa* Pol or *Mtu* PolDom protein where indicated. Incubated at 37°C for 1 hour and electrophoresed on a 15% denaturing polyacrylamide gel.

S10. *Mpa* PE can rescue NHEJ intermediates formed by *Mtu* PolDom and *Mpa* Pol.

Gap-filling, phosphoesterase and ligation reaction mixtures contained 50mM Tris pH 7.5, 5mM MnCl₂, 250μM NTPs, 30nM 5'-fluorescein labelled 16mer DNA, and 300nM *Mpa* Pol or *Mtu* PolDom, 300nM *Mpa* Lig and PE protein where indicated. Incubated at 37°C for 1 hour and electrophoresed on a 15% denaturing polyacrylamide gel.

S11. *Mpa* Pol gap filling and strand displacement at nicks is template dependent.

Polymerase gap filling and strand displacement reaction mixtures contained 50mM Tris pH 7.5, 5mM MnCl₂, 62.5μM of the indicated NTP, 30nM 5'-fluorescein labelled 16mer DNA (A, B) and 47mer DNA (C), and 300nM *Mpa* Pol protein. Incubated at 37°C for 1 hour 30 minutes and electrophoresed on a 15% denaturing polyacrylamide gel.

S11A and S11B demonstrate template dependent incorporation of NTPs onto a substrate with a downstream strand. Although *Mpa* Pol is capable of incorporating NTPs to a DNA primer in a template-independent manner (S8), these assays indicate that base selection is performed primarily by template dependence, even in the presence of an annealed sequence of downstream DNA.

S11C demonstrates the template dependent incorporation of NTPs onto a substrate with a self-annealing microhomology of 4 base pairs, which creates a single nucleotide gap on either side. Proficiently incorporation of UTP required prior formation of the microhomology, since *Mpa* Pol can only extend from a 3'-hydroxyl and template-independent incorporation function is limited to a single addition onto a double-stranded blunt DNA termini. Furthermore, preferential incorporation of UTP indicates strong template dependence.

S12. *Mpa* PE resection of strand displacement intermediates during 3' overhang NHEJ repair.

Gap-filling, phosphoesterase and ligation reaction mixtures contained 50mM Tris pH 7.5, 5mM MnCl₂, 250μM NTPs, 30nM 5'-fluorescein labelled 42mer DNA (A) and 48mer DNA (B), and 300nM *Mpa* Pol, Lig and PE and 400nM Ku protein where indicated. *Mpa* Pol pre-incubated with the substrate for 30 minutes where indicated before addition of *Mpa* Lig, PE and Ku. Electrophoresed on a 15% denaturing polyacrylamide gel.

Supporting information:

Ribonucleolytic resection is required for repair of strand displaced NHEJ intermediates, Edward Bartlett, Nigel C. Brissett and Aidan J. Doherty

S1A *Mpa* Ligase alignment

<i>Mpa</i> Lig	K	M	E	Q	-	-	-	P	E	W	E	E	P	M	L	A	T	L	T	-	S	K	R	F	D	D	H	G	W	A	F	G	H	K	L	D	G	A	R	C	L	A	Y	R	K
<i>Mtu</i> LigD	K	D	Q	K	V	F	E	F	D	N	L	A	P	M	L	A	T	H	G	T	V	A	G	L	K	A	S	Q	W	A	F	E	G	K	W	D	G	Y	R	L	L	V	E	A	D
<i>Msm</i> LigD	K	N	A	R	-	-	-	P	E	D	F	A	P	M	L	A	T	E	G	S	V	A	K	Y	K	A	K	Q	W	A	F	E	G	K	W	D	G	Y	R	V	I	I	D	A	D
<i>Pae</i> LigD	K	T	P	M	-	-	-	P	E	W	I	A	P	E	L	A	S	L	-	-	V	E	Q	P	P	H	G	Q	W	A	Y	E	L	K	L	D	G	Y	R	L	M	S	R	I	E

<i>Mpa</i> Lig	D	G	H	V	R	L	M	S	R	T	R	H	E	M	N	A	S	Y	P	E	V	A	D	A	L	M	A	Q	K	A	N	D	F	I	I	D	G	E	I	V	A	L	D	E	N
<i>Mtu</i> LigD	H	G	A	V	R	L	R	S	R	S	G	R	D	V	T	A	E	Y	P	Q	L	R	A	L	A	E	D	L	A	D	H	H	V	V	L	D	G	E	A	V	V	L	D	S	S
<i>Msm</i> LigD	H	G	Q	L	Q	I	R	S	R	T	G	R	E	V	T	G	E	Y	P	Q	F	K	A	L	A	A	D	L	A	E	H	H	V	V	L	D	G	E	A	V	A	L	D	E	S
<i>Pae</i> LigD	D	G	H	V	R	L	L	T	R	N	G	H	D	W	S	E	R	L	P	H	L	K	K	A	L	A	G	L	G	L	E	R	S	W	L	D	G	E	L	V	V	L	D	E	E

<i>Mpa</i> Lig	G	N	S	K	F	E	L	L	Q	P	R	M	G	L	E	D	P	A	E	A	R	R	T	G	I	P	V	H	Y	Y	V	F	D	I	L	F	Y	D	G	Y	E	L	T	R	L
<i>Mtu</i> LigD	G	V	P	S	F	S	Q	M	Q	N	R	G	-	-	-	-	-	-	-	-	R	D	T	R	V	E	F	W	A	F	D	L	L	Y	L	D	G	R	A	L	L	G	T		
<i>Msm</i> LigD	G	V	P	S	F	G	Q	M	Q	N	R	A	-	-	-	-	-	-	-	-	R	S	T	R	V	E	F	W	A	F	D	I	L	W	L	D	G	R	S	L	L	R	A		
<i>Pae</i> LigD	G	R	P	D	F	Q	A	L	Q	N	A	F	E	-	-	-	-	-	-	-	E	G	H	G	E	N	I	L	Y	F	L	F	D	L	P	Y	H	E	G	E	D	L	R	D	V

<i>Mpa</i> Lig	P	L	L	E	R	K	H	I	L	E	K	T	L	D	Y	G	-	-	D	T	V	R	F	L	E	D	R	L	G	G	G	L	G	L	F	D	E	A	C	K	K	G	W	E	G
<i>Mtu</i> LigD	R	Y	Q	D	R	R	K	L	L	E	T	L	A	N	A	T	-	-	-	S	L	T	V	P	E	L	L	P	G	D	G	A	Q	A	F	A	C	S	R	K	H	G	W	E	G
<i>Msm</i> LigD	K	Y	S	D	R	R	K	I	L	E	A	L	A	D	G	G	-	-	-	G	L	I	V	P	D	Q	L	P	G	D	G	P	E	A	M	E	H	V	R	K	K	R	F	E	G
<i>Pae</i> LigD	A	L	E	E	R	R	A	R	L	E	A	L	L	D	G	Q	D	E	D	P	L	R	F	S	A	T	L	A	E	N	P	R	D	L	L	A	S	A	C	R	L	G	L	E	G

<i>Mpa</i> Lig	L	I	A	K	R	A	Y	G	P	Y	V	H	-	G	R	S	D	E	W	L	K	F	K	C	V	Q	N	Q	E	F	V	I	G	G	Y	T	E	P	H	G	K	R	I	-	E
<i>Mtu</i> LigD	V	I	A	K	R	R	D	S	R	Y	Q	P	G	R	R	C	A	S	W	V	K	D	K	H	W	N	T	Q	E	V	V	I	G	G	W	R	A	G	E	G	R	S	S	G	
<i>Msm</i> LigD	V	V	A	K	K	W	D	S	T	Y	Q	P	G	R	R	S	S	S	W	I	K	D	K	I	W	N	T	Q	E	V	V	I	G	G	W	R	Q	G	E	G	R	S	S	G	
<i>Pae</i> LigD	V	I	G	K	R	L	G	S	V	Y	R	S	-	R	R	S	N	D	W	I	K	L	K	C	Q	L	R	Q	E	F	V	I	V	G	Y	T	E	P	K	G	S	R	R	-	H

<i>Mpa</i> Lig	F	G	A	I	L	V	G	Y	Y	Q	D	-	-	-	G	K	L	M	Y	A	G	K	V	G	T	G	F	D	E	K	T	L	R	S	L	M	D	K	F	R	P	L	E	R	K
<i>Mtu</i> LigD	V	G	S	L	L	M	G	I	P	G	P	-	-	-	G	G	L	Q	F	A	G	R	V	G	T	G	L	S	E	R	E	L	A	N	L	K	E	M	L	A	P	L	H	T	D
<i>Msm</i> LigD	I	G	A	L	V	L	G	I	P	G	P	-	-	-	E	G	L	Q	F	V	G	R	V	G	T	G	F	T	E	K	E	L	S	K	L	K	D	M	L	K	P	L	H	T	D
<i>Pae</i> LigD	I	G	A	L	L	L	G	L	Y	S	P	D	E	E	R	R	L	R	Y	A	G	K	V	G	T	G	F	T	A	A	S	L	K	K	V	H	E	R	L	A	P	L	A	I	N

<i>Mpa</i> Lig	A	S	P	F	A	E	E	-	-	-	V	R	E	K	E	V	H	W	L	E	P	K	L	V	A	E	I	G	F	E	W	T	D	Y	G	K	L	R	Q	P	R	Y	D	G		
<i>Mtu</i> LigD	E	S	P	F	D	V	P	L	P	A	R	D	A	K	G	I	T	Y	V	K	P	A	L	V	A	E	V	R	Y	S	E	W	T	P	E	G	R	L	R	Q	P	S	S	W	R	G
<i>Msm</i> LigD	E	S	P	F	N	A	P	L	P	K	V	D	A	R	G	V	T	F	V	R	P	E	L	V	G	E	V	R	Y	S	E	R	T	S	D	G	R	L	R	Q	P	S	S	W	R	G
<i>Pae</i> LigD	S	S	P	L	A	K	V	P	P	A	R	E	T	G	V	V	Q	W	V	R	P	E	R	L	C	E	V	S	Y	A	Q	M	T	R	G	G	I	I	R	Q	A	V	F	H	G	

<i>Mpa</i> Lig	L	R	Y	D	K	D	P	H	S	V	I	R	E	V	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>Mtu</i> LigD	L	R	P	D	K	K	P	S	E	V	V	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
<i>Msm</i> LigD	L	R	P	D	K	T	P	D	E	V	V	W	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
<i>Pae</i> LigD	L	R	E	D	K	P	A	R	E	V	T	G	E	R	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		

S1. A, B, C & D Protein sequence alignments of *Mpa* Lig, Pol, PE and Ku with *Mtu*, *Msm*, *Pae* LigD and Ku.

(A) *Methanocella paludicola* (*Mpa*) Ligase protein sequence alignment with Ligase D (LigD) ligase domains (LigDom) of *Mycobacterium tuberculosis*

(*Mtu*), *Mycobacterium Smegmatis* (*Msm*) and *Pseudomonas Aeruginosa* (*Pae*). **B.** *Mpa* polymerase (Pol) protein sequence alignment with LigD polymerase domains (PolDom) of *Mtu*, *Msm* and *Pae*. **C.** *Mpa* phosphoesterase (PE) protein sequence alignment with LigD phosphoesterase domains (PEDom) of *Mtu*, *Msm* and *Pae*. **D.** *Mpa* Ku protein sequence alignment with Ku of *Mtu*, *Msm* and *Pae*. All alignments were made with TCoffee and edited in JalView.

S1B *Mpa* Pol alignment

<i>Mpa</i> Pol	H D I K V T N P D K V L F P E D G I T K G E L V D Y Y R R I S G V M V P L V R G R P
<i>Mtu</i> LigD	Q R V T L T N A D K V L Y P A T G T T K S D I F D Y Y A G V A E V M L G H I A G R P
<i>Msm</i> LigD	E R V R L T N P D K V L Y P A T G T T K A E V F D Y Y L S I A Q V M V P H I A G R P
<i>Pae</i> LigD	A G V R I S H P R R L I D P S I Q A S K L E L A E F H A R Y A D L L L R D L R E R P

<i>Mpa</i> Pol	M T M Q R F P D G I G K E G F F Q K E A S D Y - F P D W V H R - A T L E L G K G G I
<i>Mtu</i> LigD	A T R K R W P N G V D Q P A F F E K Q L A L S - A P P W L S R - A T V A H R S G T T
<i>Msm</i> LigD	V T R K R W P N G V A E E A F F E K Q L A S S - A P S W L E R - G S I T H K S G T T
<i>Pae</i> LigD	V S L V R G P D G I G G E L F F Q K H A A R L K I P G I V Q L D P A L D P G H P P -

<i>Mpa</i> Pol	Q H Q V V C D D A A T L V Y L A S Q A M I T P H V F L S R I D K - - - - - V - - H
<i>Mtu</i> LigD	T Y P - I I D S A T G L A W I A Q Q A A L E V H V P Q W R F V A E P G S G E L N P G
<i>Msm</i> LigD	T Y P - I I N T R E G L A W V A Q Q A S L E V H V P Q W R F E D - - - - - G D Q G
<i>Pae</i> LigD	- - L L Q I R S A E A L V G A V Q M G S I E F H T W N A S L A N - - - - - L - - E

<i>Mpa</i> Pol	Y P D R L I F D L D P P - D N N F E T - V R S A A K T I R E A L D A E G Y P V Y L M T
<i>Mtu</i> LigD	P A T R L V F D L D P G - E G V M M A Q L A E V A R A V R D L L A D I G L V T F P V T
<i>Msm</i> LigD	P A T R I V F D L D P G - E G V T M T Q L C E I A H E V R A L M T D L D L E T Y P L T
<i>Pae</i> LigD	R P D R F V L D L D P D P A L P W K R - M L E A T Q L S L T L L D E L G L R A F L K T

<i>Mpa</i> Pol	T G S R G L H V V V P L D R S A D F D T V R A F A R G F G E K L T K K Y P D R F T I
<i>Mtu</i> LigD	S G S K G L H L Y T P L D E P V S S R G A T V L A K R V A Q R L E Q A M P A L V T S
<i>Msm</i> LigD	S G S K G L H L Y V P L A E P I S S R G A S V L A R R V A Q Q L E Q A M P K L V T A
<i>Pae</i> LigD	S G G K G M H L L V P L E R R H G W D E V K A F A Q A I S Q H L A R L M P E R F S A

<i>Mpa</i> Pol	E L S K E K R R G R L F L D Y L R N S Y G Q T G V A P Y G V R A R S G A P V A T P I T
<i>Mtu</i> LigD	T M T K S L R A G K V F V D W S Q N S G S K T T I A P Y S L R G R T H P T V A A P R T
<i>Msm</i> LigD	A M T K S L R A G K V F L D W S Q N N A A K T T I A P Y S L R G R D H P T V A A P R T
<i>Pae</i> LigD	V S G P R N R V G R I F V D Y L R N S R G A S T V A A Y S V R A R E G L P V S V P V F

<i>Mpa</i> Pol	W D E L D D I S G S Q E Y N I R N I M G R M D K R G - - D A W K Y I D
<i>Mtu</i> LigD	W A E L D D P A L R - Q L S Y D E V I L T R I A R D G - - D I L L E R L D
<i>Msm</i> LigD	W D E I A D P E L R - H L R F D E V I L D R L D E Y G - - D I L L A P L D
<i>Pae</i> LigD	R E E L D S L Q G A N Q W N L R S L P Q R L D E L A G D D P W A D Y A

S1C Mpa PE alignment

<i>Mpa</i> PE	E	P	L	E	S	Y	R	S	K	R	D	F	E	K	T	P	E	P	A	G	G	E	E	K	E	-	I	K	H	P	-	I	Y	V	I	Q	K	H	D	A	S	H	L
<i>Mtu</i> LigD	D	R	L	T	R	Y	R	R	M	R	D	A	S	K	T	P	E	P	I	P	T	A	K	P	V	T	G	D	G	N	-	T	F	V	I	Q	E	H	H	A	R	R	P
<i>Msm</i> LigD	D	K	L	T	T	Y	R	S	M	R	D	A	S	K	T	P	E	P	V	P	K	E	I	P	K	T	G	N	N	D	-	K	F	V	I	Q	E	H	H	A	R	R	L
<i>Pae</i> LigD	K	P	L	A	E	Y	A	R	K	R	D	F	R	Q	T	P	E	P	S	G	R	K	R	R	K	D	G	D	G	L	L	R	Y	C	I	Q	K	H	D	A	S	R	L

<i>Mpa</i> PE	H	Y	D	L	R	L	E	M	G	G	V	L	K	S	W	A	V	P	K	G	P	S	L	D	P	K	V	K	R	L	A	M	P	T	E	D	H	P	I	G	Y	A	T
<i>Mtu</i> LigD	H	Y	D	F	R	L	E	C	D	G	V	L	V	S	W	A	V	P	K	N	L	P	D	N	T	S	V	N	H	L	A	I	H	T	E	D	H	P	L	E	Y	A	T
<i>Msm</i> LigD	H	Y	D	L	R	L	E	R	D	G	V	L	V	S	F	A	V	P	K	N	L	P	E	T	T	A	E	N	R	L	A	V	H	T	E	D	H	P	I	E	Y	L	A
<i>Pae</i> LigD	H	Y	D	F	R	L	E	L	D	G	T	L	K	S	W	A	V	P	K	G	P	C	L	D	P	A	V	K	R	L	A	V	Q	V	E	D	H	P	L	D	Y	A	D

<i>Mpa</i> PE	F	E	G	V	I	P	E	G	Q	Y	G	G	T	V	M	V	W	D	I	G	T	Y	R	N	L	R	E	E	K	P	E	G	S	R	M	T	I	E	Q	S	Y	D	
<i>Mtu</i> LigD	F	E	G	A	I	P	S	G	E	Y	G	A	G	K	V	I	I	W	D	S	G	T	Y	D	T	E	K	F	H	-	-	-	-	-	-	-	-	D	-	-	D	P	H
<i>Msm</i> LigD	F	H	G	S	I	P	K	G	E	Y	G	A	G	D	M	V	I	W	D	S	G	S	Y	E	T	E	K	F	R	V	P	E	E	L	-	-	-	D	N	P	D	D	S
<i>Pae</i> LigD	F	E	G	S	I	P	K	G	H	Y	G	A	G	D	V	I	V	W	D	R	G	A	W	T	P	L	D	D	-	-	-	-	-	-	-	-	-	P	R	E	G	L	E

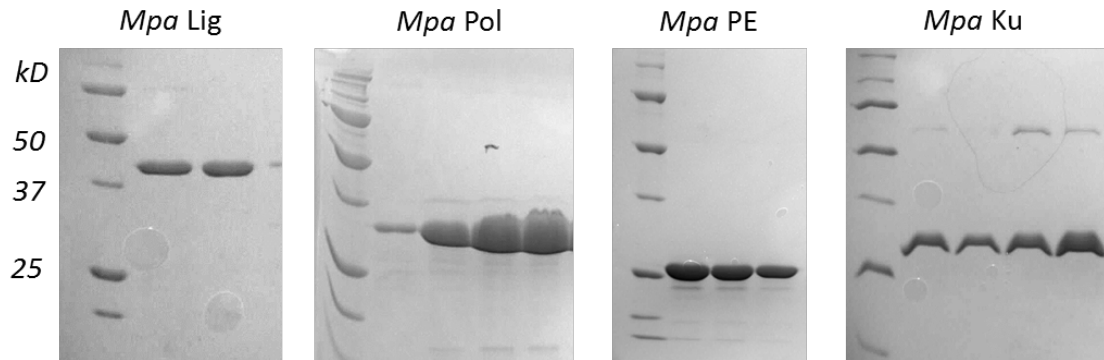
<i>Mpa</i> PE	Q	G	K	I	E	V	F	L	E	G	K	K	L	K	G	S	Y	A	L	I	R	T	G	G	I	-	E	K	R	G	W	L	F	F	K	M	K	E	P	H	E	G	S
<i>Mtu</i> LigD	T	G	E	V	I	V	N	L	H	G	G	R	I	S	G	R	Y	A	L	I	R	T	N	G	-	-	-	-	D	R	W	L	A	H	R	L	K	N	Q	K	D	Q	K
<i>Msm</i> LigD	H	G	E	I	I	V	T	L	H	G	E	K	V	D	G	R	Y	A	L	I	Q	T	K	G	-	-	-	-	K	N	W	L	A	H	R	M	K	D	Q	K	N	A	R
<i>Pae</i> LigD	K	G	H	L	S	F	L	D	G	E	K	L	S	G	R	W	H	L	I	R	T	N	L	R	G	K	Q	P	Q	W	F	L	V	K	S	K	D	D	E	A	R	S	

<i>Mpa</i> PE	Y	E	D	I	E	K	-	-	A	A	P	D	S	V	L
<i>Mtu</i> LigD	V	F	E	F	-	-	-	-	-	-	-	-	-	-	-
<i>Msm</i> LigD	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>Pae</i> LigD	L	D	R	F	D	V	L	K	E	R	P	D	S	V	L

S1D Mpa Ku alignment

<i>Mpa</i> Ku	M	-	Q	A	I	W	S	G	N	L	A	F	G	T	I	L	I	P	V	R	M	Y	A	A	S	E	E	L	R	V	P	F	H	L	V	H	A	E	D	G	G	
<i>Mtu</i> Ku	M	-	R	A	I	W	T	G	S	I	A	F	G	L	V	N	V	P	V	K	V	Y	S	A	T	A	D	H	D	I	R	F	H	Q	V	H	A	K	D	N	G	
<i>Msm</i> Ku	M	-	R	S	I	W	K	G	S	I	A	F	G	L	V	N	V	P	V	K	V	Y	S	A	T	E	D	H	D	I	K	F	H	Q	V	H	A	K	D	N	G	
<i>Pae</i> Ku	M	A	R	A	I	W	K	G	A	I	S	F	G	L	V	H	I	P	V	S	L	S	A	A	A	S	S	Q	G	I	D	F	D	W	L	D	Q	R	S	M	D	
<i>Mpa</i> Ku	R	V	K	Y	K	K	V	C	E	L	D	G	K	E	L	H	M	S	D	I	H	K	G	Y	E	I	G	D	-	K	M	L	Q	F	T	D	E	E	L	D	E	
<i>Mtu</i> Ku	R	I	R	Y	K	R	V	C	E	A	C	G	E	V	V	D	Y	R	D	L	A	R	A	Y	E	S	G	D	G	Q	M	V	A	I	T	D	D	D	I	A	S	
<i>Msm</i> Ku	R	I	R	Y	K	R	V	C	E	V	C	G	E	V	V	E	Y	R	D	I	N	K	A	F	E	S	D	D	G	Q	M	V	V	I	T	D	E	D	I	S	T	
<i>Pae</i> Ku	P	V	G	Y	K	R	V	N	K	V	T	G	K	E	I	E	R	E	H	I	V	K	G	V	E	Y	E	K	G	R	Y	V	V	L	S	E	E	E	I	R	A	
<i>Mpa</i> Ku	L	R	P	V	L	S	K	S	L	K	I	L	G	F	C	E	Q	E	E	V	P	L	L	A	L	E	Q	P	F	Y	L	G	T	E	S	R	K	K	G	V	A	
<i>Mtu</i> Ku	L	P	E	E	R	S	R	E	I	E	V	L	E	F	V	P	A	A	D	V	D	P	M	M	F	D	R	S	Y	F	L	E	P	D	S	K	S	-	-	-	S	
<i>Msm</i> Ku	L	P	E	E	R	S	R	E	I	E	V	L	E	F	I	P	A	E	Q	L	D	P	L	M	Y	D	K	S	Y	F	L	E	P	D	S	K	S	-	-	-	S	
<i>Pae</i> Ku	A	H	P	K	S	T	Q	T	I	E	I	F	A	F	V	D	S	Q	E	I	P	L	Q	H	F	D	T	P	Y	Y	L	S	P	D	R	R	G	-	-	-	G	
<i>Mpa</i> Ku	Q	P	F	A	L	V	K	K	A	M	E	K	S	G	K	V	A	I	V	S	W	V	T	R	A	S	E	Q	I	G	M	M	M	P	Y	-	-	-	-	E	K	
<i>Mtu</i> Ku	K	S	Y	V	L	L	A	K	T	L	A	E	T	D	R	M	A	I	V	H	F	T	L	R	N	K	T	R	L	A	A	L	R	V	K	D	F	G	K	R	E	
<i>Msm</i> Ku	K	S	Y	V	L	L	A	K	T	L	A	E	T	D	R	I	A	I	V	H	F	S	L	R	N	K	S	R	L	A	A	L	R	V	K	D	F	S	K	R	D	
<i>Pae</i> Ku	K	V	Y	A	L	L	R	E	T	L	E	S	T	G	K	V	A	L	A	N	V	V	L	H	T	R	Q	H	L	A	L	L	R	P	L	-	-	-	-	E	D	
<i>Mpa</i> Ku	G	F	L	I	K	A	I	L	Y	Q	Q	Q	I	R	P	F	D	R	V	E	V	Q	E	S	-	-	-	-	E	V	S	E	D	L	I	D	K	G	I	M	L	
<i>Mtu</i> Ku	V	M	M	V	H	T	L	L	W	P	D	E	I	R	D	P	D	F	P	V	L	D	Q	K	V	-	-	-	E	I	K	P	A	E	L	K	M	A	G	Q	V	
<i>Msm</i> Ku	V	M	M	I	H	T	L	L	W	P	D	E	I	R	D	P	D	F	P	I	L	D	K	E	V	-	-	-	Q	I	K	P	A	E	L	K	M	A	G	Q	V	
<i>Pae</i> Ku	A	L	V	L	I	T	L	R	W	P	S	Q	V	R	S	L	D	G	L	E	L	D	E	S	V	T	A	A	K	L	D	R	R	E	L	E	M	A	R	R	L	
<i>Mpa</i> Ku	I	E	R	M	T	F	K	F	D	H	A	A	Y	K	E	S	Y	S	Q	A	L	R	E	I	I	E	A	K	A	L	G	K	E	V	-	K	V	E	P	V	-	
<i>Mtu</i> Ku	V	D	S	M	A	D	D	F	N	P	D	R	Y	H	D	T	Y	Q	E	Q	L	Q	E	L	I	D	T	K	L	E	G	G	Q	A	F	T	A	E	D	Q	P	
<i>Msm</i> Ku	V	E	S	M	T	D	D	F	K	P	D	L	Y	H	D	D	Y	Q	E	Q	L	R	E	L	V	Q	A	K	L	E	G	G	E	A	F	S	V	E	E	Q	P	
<i>Pae</i> Ku	V	E	D	M	A	S	H	W	Q	P	D	E	Y	K	D	S	F	S	D	K	I	M	K	L	V	E	E	K	A	A	K	G	Q	L	H	A	V	E	E	E	E	
<i>Mpa</i> Ku	-	-	-	V	K	S	-	A	E	T	R	S	L	E	A	E	L	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
<i>Mtu</i> Ku	R	L	L	D	E	P	-	E	D	V	S	D	L	L	A	K	L	E	A	S	V	K	A	R	S	K	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
<i>Msm</i> Ku	A	E	L	D	E	G	T	E	D	V	S	D	L	L	A	K	L	E	A	S	V	K	A	R	K	G	G	K	S	-	-	-	-	-	-	-	-	-	-	-	-	
<i>Pae</i> Ku	E	V	A	G	K	G	-	A	D	I	I	D	L	T	D	L	L	K	R	S	L	R	P	R	A	G	G	G	K	-	-	-	-	-	-	-	-	-	-	-	-	-

S2 *Mpa* NHEJ enzymes SDS-PAGE



S2. SDS-PAGE *Mpa* Lig, Pol, PE and Ku.

SDS-PAGE of gel filtration chromatography fractions of respective purifications of *Mpa* Lig, Pol, PE and Ku. Migration sizes of the proteins were: *Mpa* Lig ~40kD, *Mpa* Pol ~34kD, *Mpa* PE ~24kD and *Mpa* Ku ~30kD (monomer) ~60kD (dimer). The gels indicate the high purity of the proteins.

S3

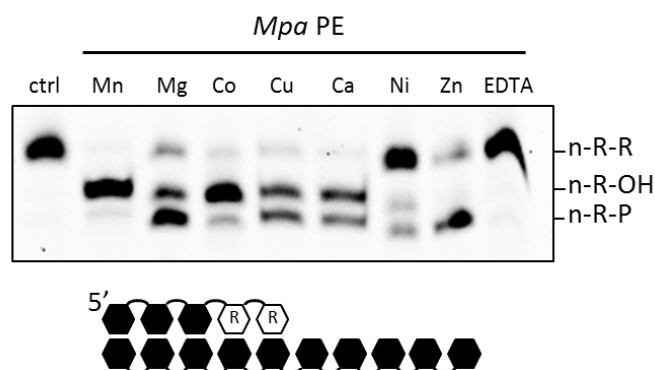
Archaeo-eukaryotic primase (AEP) superfamily motifs

	Motif I	Motif II	Motif III	
<i>M. paludicola</i>	R L I F D L D P P D	T G S R G L H V V V	R L F L D Y	NHEJ Pol
<i>M. tuberculosis</i>	R L V F D L D P G E	S G S K G L H L Y T	K V F V D W	
<i>P. aeruginosa</i>	R F V L D L D P D P	S G G K G M H L L V	K I F V D Y	
<i>H. sapiens</i>	E L V F D I D M T D	S G R R G V H C W V	F P R L D I	Primase
<i>M. musculus</i>	E L V F D I D M T D	S G R R G V H C W V	F P R L D V	
<i>D. melanogaster</i>	E L V F D I D M T D	S G R R G I H C W V	Y P R L D I	
<i>S. cerevisiae</i>	E L V F D I D M D D	S G R R G A H C W V	Y P K L D V	
<i>S. pombe</i>	E L V F D I D M T D	S G R R G I H A W I	Y P R L D V	

S3. Archaeo-eukaryotic primase (AEP) superfamily motif sequence alignment.

Three key conserved motifs (I, II and III) confer membership to the AEP superfamily. Here *Mpa* Pol is aligned with *Mycobacterium tuberculosis* and *Pseudomonas Aeruginosa* NHEJ Pols, and *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* primase small sub-units.

S4

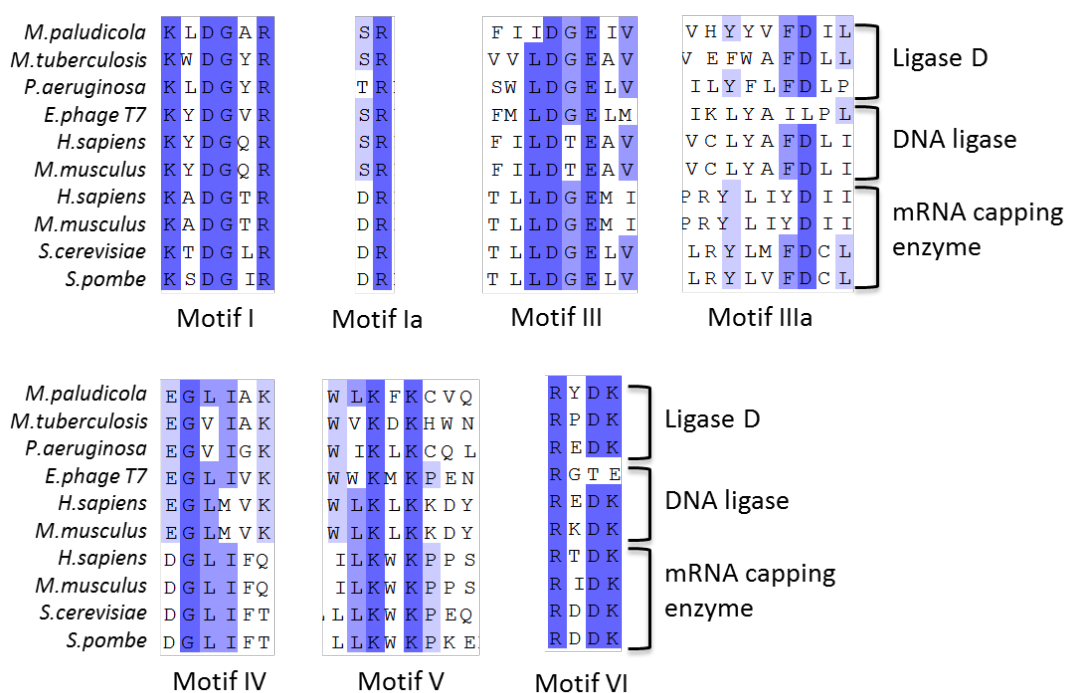


S4. *Mpa* PE ion preference.

Phosphoesterase reaction mixtures contained 50mM Tris pH7.5, 5mM MnCl₂, MgCl₂, CoCl₂, CuCl₂, CaCl₂, NiCl₂ or Zn(O₂CCH₃)₂ where indicated, 30nM 5'-fluorescein labelled 16mer DNA/RNA (D14R2), and 300nM *Mpa* PE protein. Incubated at 37°C for 1 hour 30 minutes and electrophoresed on a 15% denaturing polyacrylamide gel.

S5

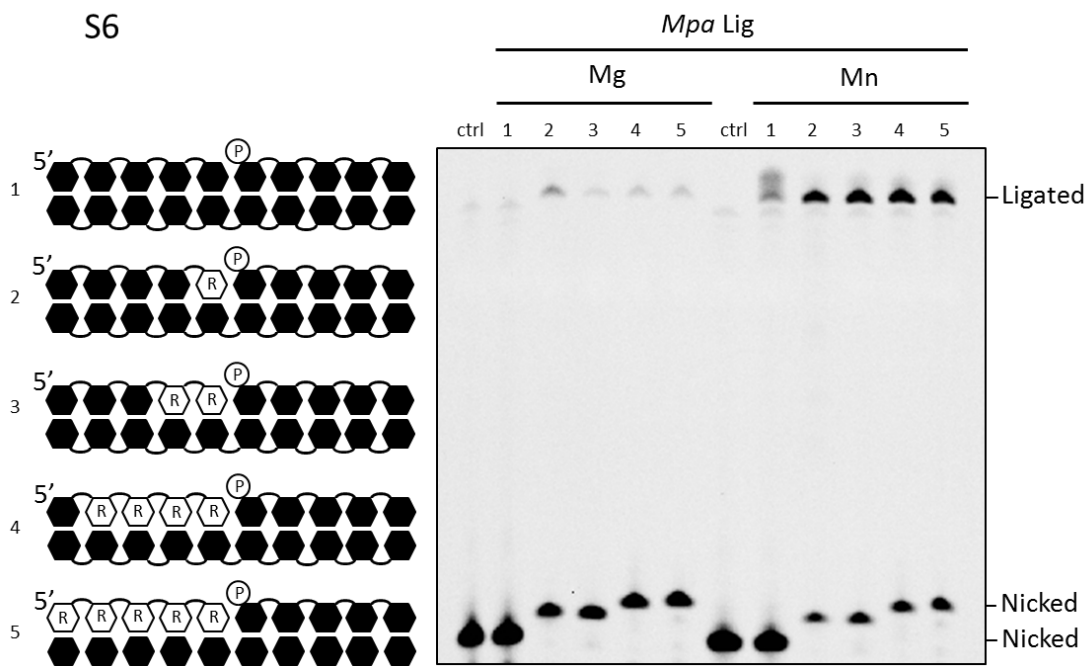
Polynucleotide ligase and RNA capping superfamily motifs



S5. Polynucleotide ligase and mRNA capping superfamily motifs

Six key conserved motifs (I, Ia, III, IIIa, IV and V) confer membership to the polynucleotide ligase and mRNA capping superfamily. Motif VI is also included with highly conserved residues. Here *Mpa* Lig is aligned with *Mycobacterium tuberculosis* and *Pseudomonas aeruginosa* Ligase D; *Enterobacterio phage T7*, *Homo sapiens*, *Mus musculus* DNA ligase; and *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* mRNA capping enzymes.

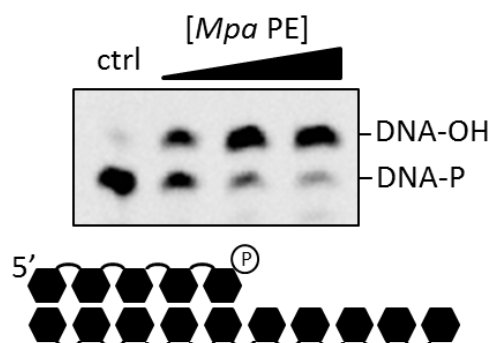
S6



S6. *Mpa* Lig preferable ligation with Mn, not Mg

Ligase reaction mixtures contained 50mM Tris pH7.5, 5mM MnCl₂ or Mg Cl₂, 30nM 5'-fluorescein labelled 16mer oligomer (D16, D15R1, D14R2, D12R4 or D8R8 where indicated) and 19mer 5'-phosphorylated DNA, and 300nM *Mpa* Lig protein. Incubated at 37°C for 2 hours and electrophoresed on a 15% denaturing polyacrylamide gel.

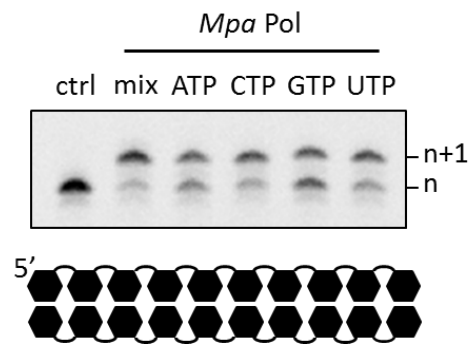
S7



S7. *Mpa* PE phosphomonoesterase activity does not require 2'-OH

Phosphatase reaction mixtures contained 50mM Tris pH7.5, 5mM MnCl₂, 30nM 5'-fluorescein labelled 16mer DNA with 3'-phosphate, and 250, 500 and 1000nM *Mpa* PE protein. Incubated at 37°C for 1 hour 30 minutes and electrophoresed on a 15% denaturing polyacrylamide gel.

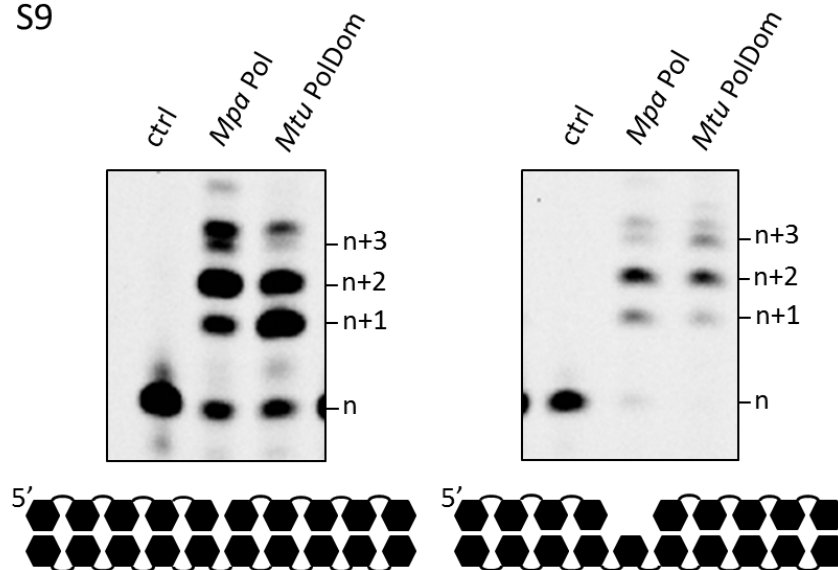
S8



S8. *Mpa* Pol non-templated NTP incorporation

Polymerase non-templated extension reaction mixtures contained 50mM Tris pH7.5, 5mM MnCl_2 , $62.5\mu\text{M}$ ATP, CTP, GTP, UTP or $250\mu\text{M}$ mixed NTPs where indicated, 30nM 5'-fluorescein labelled 36mer DNA, and 300nM *Mpa* Pol protein. Incubated at 37°C for 1 hour and electrophoresed on a 15% denaturing polyacrylamide gel.

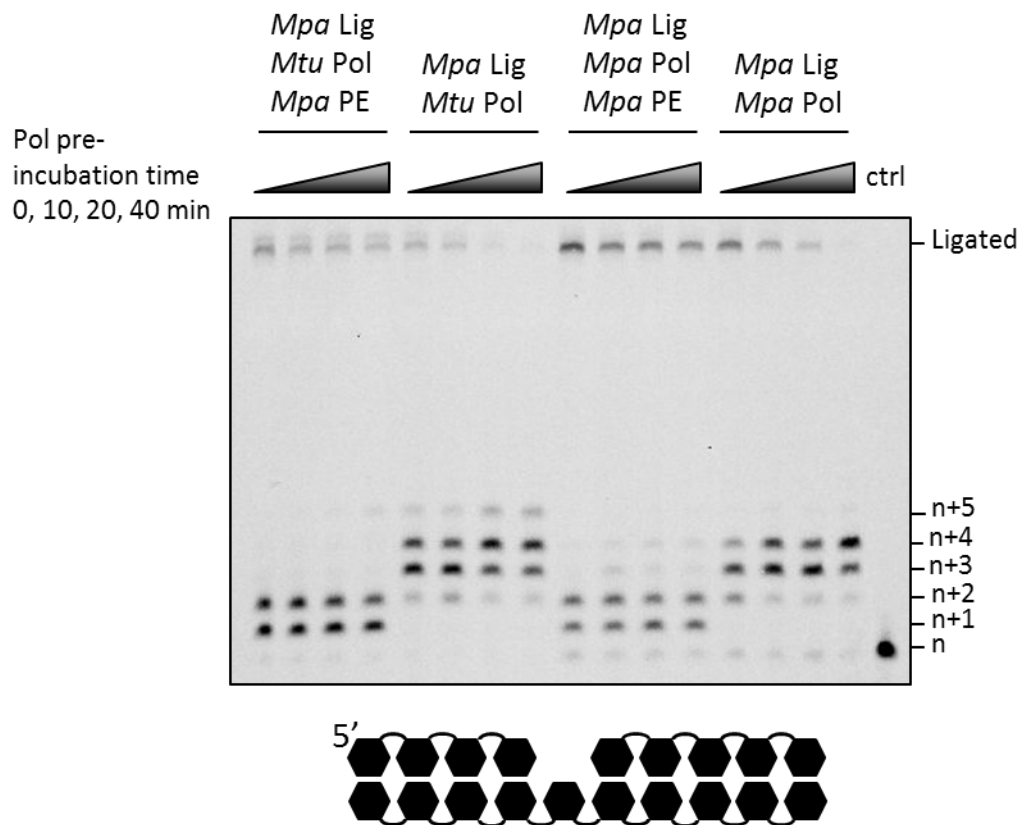
S9



S9. *Mtu* PolDom strand displacement

Polymerase strand displacement reaction mixtures contained 50mM Tris pH7.5, 5mM MnCl_2 , $250\mu\text{M}$ NTPs, 30nM 5'-fluorescein labelled 16mer DNA, and 300nM *Mpa* Pol or *Mtu* PolDom protein where indicated. Incubated at 37°C for 1 hour and electrophoresed on a 15% denaturing polyacrylamide gel.

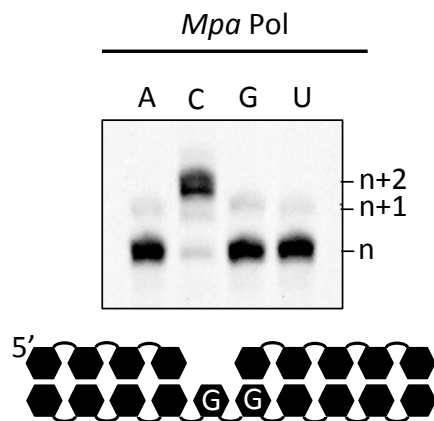
S10



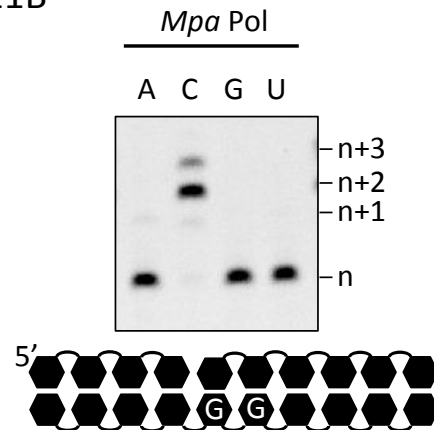
S10. *Mpa* PE can rescue NHEJ intermediates formed by *Mtu* PolDom and *Mpa* Pol.

Gap-filling, phosphoesterase and ligation reaction mixtures contained 50mM Tris pH7.5, 5mM MnCl₂, 250μM NTPs, 30nM 5'-fluorescein labelled 16mer DNA, and 300nM *Mpa* Pol or *Mtu* PolDom, 300nM *Mpa* Lig and PE protein where indicated. Incubated at 37°C for 1 hour and electrophoresed on a 15% denaturing polyacrylamide gel.

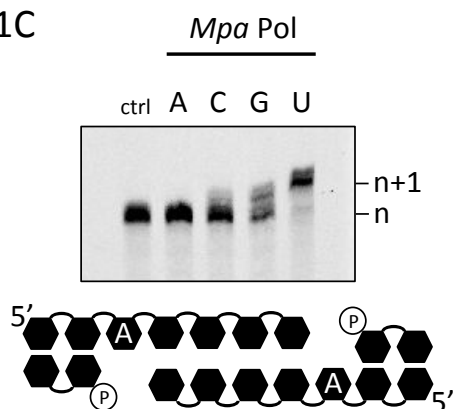
S11A



S11B



S11C



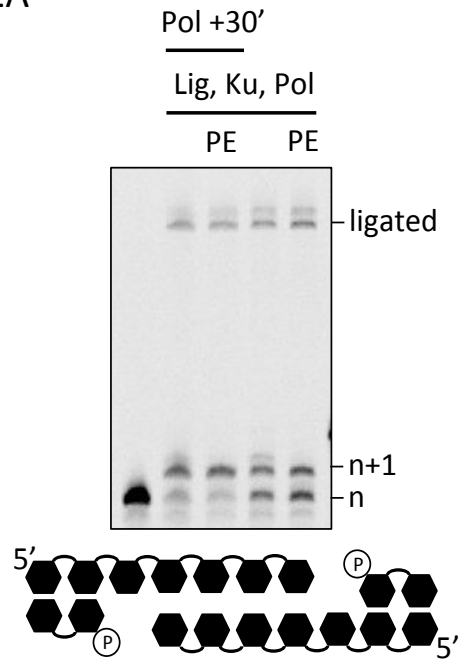
S11. *Mpa Pol* gap filling and strand displacement at nicks is template dependent.

Polymerase gap filling and strand displacement reaction mixtures contained 50mM Tris pH7.5, 5mM $MnCl_2$, 62.5 μ M of the indicated NTP, 30nM 5'-fluorescein labelled 16mer DNA (A, B) and 47mer DNA (C), and 300nM *Mpa Pol* protein. Incubated at 37°C for 1 hour 30 minutes and electrophoresed on a 15% denaturing polyacrylamide gel.

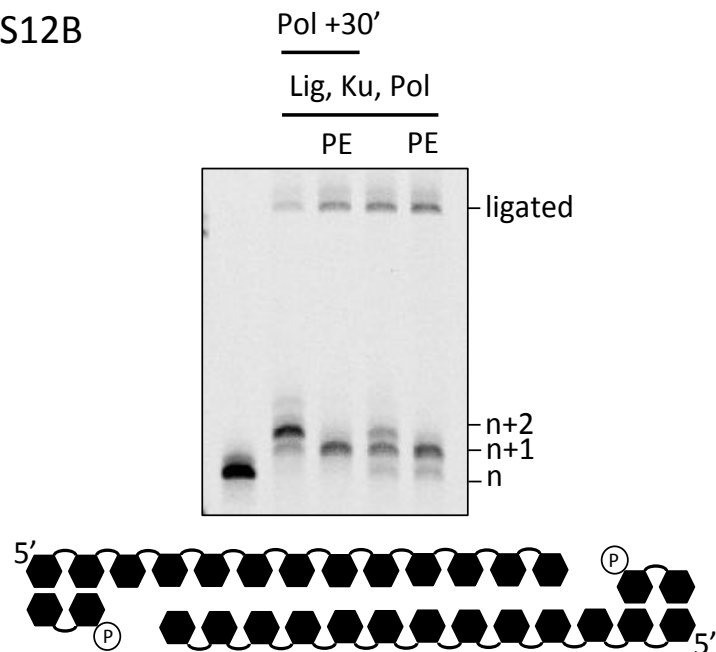
S11A and S11B demonstrate template dependent incorporation of NTPs onto a substrate with a downstream strand. Although *Mpa Pol* is capable of incorporating NTPs to a DNA primer in a template-independent manner (S8), these assays indicate that base selection is performed primarily by template dependence, even in the presence of an annealed sequence of downstream DNA.

S11C demonstrates the template dependent incorporation of NTPs onto a substrate with a self-annealing microhomology of 4 base pairs, which creates a single nucleotide gap on either side. Proficient incorporation of UTP required prior formation of the microhomology, since *Mpa Pol* can only extend from a 3'-hydroxyl and template-independent incorporation function is limited to a single addition onto a double-stranded blunt DNA termini. Furthermore, preferential incorporation of UTP indicates strong template dependence.

S12A



S12B



S12. *Mpa* PE resection of strand displacement intermediates during 3' overhang NHEJ repair.

Gap-filling, phosphoesterase and ligation reaction mixtures contained 50mM Tris pH7.5, 5mM $MnCl_2$, 250 μ M NTPs, 30nM 5'-fluorescein labelled 42mer DNA (A) and 48mer DNA (B), and 300nM *Mpa* Pol, Lig and PE and 400nM Ku protein where indicated. *Mpa* Pol pre-incubated with the substrate for 30 minutes where indicated before addition of *Mpa* Lig, PE and Ku. Electrophoresed on a 15% denaturing polyacrylamide gel.